Shielding voices: The modulation of binding processes between voice features and response features by task representations

Johanna Bogon, Hedwig Eisenbarth, Steffen Landgraf & Gesine Dreisbach

To cite this article: Johanna Bogon, Hedwig Eisenbarth, Steffen Landgraf & Gesine Dreisbach (2016): Shielding voices: The modulation of binding processes between voice features and response features by task representations, *The Quarterly Journal of Experimental Psychology*, DOI: [10.1080/17470218.2016.1209686](https://doi.org/10.1080/17470218.2016.1209686)

To link to this article: <http://dx.doi.org/10.1080/17470218.2016.1209686>

Accepted author version posted online: 06 Jul 2016.
Published online: 06 Jul 2016.

Submit your article to this journal [↗](#)

View related articles [↗](#)

View Crossmark data [↗](#)

Publisher: Taylor & Francis & The Experimental Psychology Society

Journal: *The Quarterly Journal of Experimental Psychology*

DOI: 10.1080/17470218.2016.1209686

Shielding voices: The modulation of binding processes between voice features and response features by task representations

Johanna Bogon¹, Hedwig Eisenbarth², Steffen Landgraf^{3,4}, & Gesine Dreisbach¹

¹ Department of Experimental Psychology, University of Regensburg, Germany

² Department of Psychology, University of Southampton, United Kingdom

³ Department of Forensic Psychiatry and Psychotherapy, University of Regensburg, Germany

⁴ Berlin School of Mind and Brain, Humboldt University Berlin, Germany

Corresponding author: Johanna Bogon, Department of Experimental Psychology, University of Regensburg, Universitätsstraße 31, D-93053 Regensburg, Germany, Tel.: +49 (0)941 943-3814, Fax: +49 (0)941 943-1995, johanna.bogon@psychologie.uni-regensburg.de

ACCEPTED MANUSCRIPT

Abstract

During speech perception, we do not only process semantic-linguistic content but also information about the identity and the motivational and emotional state of the speaker. Furthermore, most vocal events have implications for our actions and therefore include action-related features. But the relevance and irrelevance of vocal features varies from task to task. In the present study, the authors investigate binding processes for perceptual and action-related features of spoken words and how these binding processes are modulated by the task representation of the listener. Participants had to react with two response keys to the recordings of eight different words spoken by a male or a female voice (Experiment 1) or spoken by an angry or neutral male voice (Experiment 2). There were two different instruction conditions: half of the participants learned eight stimulus-response mappings by rote (SR), and half of the participants applied a binary task rule (TR). In both experiments, SR instructed participants show clear evidence for binding processes between voice and response features indicated by an interaction between the irrelevant voice feature (speaker identity, vocal affective information) and the response. By contrast, as indicated by a three-way interaction with instruction (SR vs. TR), no such binding was found in the TR instructed groups. The results of this study are suggestive of binding and shielding as two very adaptive mechanisms that ensure successful communication and action in a dynamic social environment.

Keywords: task shielding, feature binding, partial repetition costs, voice perception, social-affective features

Introduction

The human voice is the most complex multi-feature event we are frequently confronted with. Imagine you are standing in a crowded bus near the door. Somebody in the back vocally expresses that he has to get out at the next station. Even with the person not being in your visual field, you are able to extract lots of information from his vocalization. You may realize that the speaker is a stressed-out, middle-aged male person who expects you stepping aside. This example shows that, during speech perception, we do not only process the semantic-linguistic content. The voice of the speaker offers much additional information, e.g. about the age and sex, the origin and the motivational and emotional state of the speaker. Furthermore, most vocal events have implications for our actions and therefore include action-related features. In our nervous system, these different features of spoken words are processed in a distributed manner (Belin, Fecteau, & Bédard, 2004; Jeannerod, 1999). Despite this distributed processing, normally, the different feature representations converge into a coherent perception of vocal events. In this study, we investigate binding processes that reassemble the different perceptual, affective and action-related features of spoken words. Moreover, we were interested in the flexibility and boundary conditions of such binding mechanisms. Note that the relevance and irrelevance of vocal features varies from situation to situation and from task to task. For example in an emergency call center, the people calling with the most excited voices may not be the people that are in the most urgent need of help. To make the right decisions, the physicians and assistants ideally focus on the semantic-linguistic content and their actions do not get distracted by irrelevant voice features. Thus, the second aim of this study was to investigate the flexibility of voice-related binding mechanisms, i.e. their modulation by the task representation of the listener.

It is a general organizational principle of our brain that the various features of sensory input are processed in a distributed manner (Felleman & Van Essen, 1991). According to this principle, in their model of voice perception, Belin and colleagues (2004) have proposed that after an initial processing in voice-selective areas (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000), the three main types of vocal information are further processed in three partially dissociable functional pathways: a pathway for analysis of semantic-linguistic content, a pathway for analysis of speaker identity and a pathway for vocal affective information. This

distinction has received support from neuroimaging and clinical studies (Belin, Bestelmeyer, Latinus, & Watson, 2011; Formisano, Martino, Bonte, & Goebel, 2008; Grandjean et al., 2005; Heilman, Scholes, & Watson, 1975; Kriegstein & Giraud, 2004; Schirmer, Zysset, Kotz, & Yves von Cramon, 2004; Van Lancker & Canter, 1982; Van Lancker, Kreiman, & Cummings, 1989). A prominent example is 'phonagnosia'. People with this deficit have difficulties in voice identity recognition whereas the processing of the semantic-linguistic content is preserved (Van Lancker & Canter, 1982; Van Lancker et al., 1989). With regard to vocal affective information there exists a likewise dissociation. Patients with auditory affective agnosia have difficulties judging the emotional aspects of speech while the perception of the semantic-linguistic content is unaffected (Heilman et al., 1975). Despite the distributed processing of semantic-linguistic content, speaker identity information and affective information of human vocalizations, we usually perceive spoken words as coherent events. In the visual domain, there is evidence, that binding mechanisms reassemble the different features of an object, and thus ensure its coherent perception. According to Treisman's *feature integration theory*, features belonging to the same object are bound together via an "object file", a temporary episodic representation of the object that contains the traces to the distributed feature representations (Kahneman, Treisman, & Gibbs, 1992; Treisman, 1996). Moreover, such binding processes are not restricted to perceptual stimulus features only but also to response-related features. In particular, Hommel (1998, 2004) extended the concept of "object files" and suggested so-called "event files", that establish transient bindings between perceptual as well as action-related features of an event (e.g., a stimulus-response event). One effect that is interpreted as evidence for binding processes is the occurrence of partial-repetition costs. In a simple two-choice reaction time task, partial repetition costs as a consequence of transient stimulus-response bindings can be exemplified as follows: responding to a green word with a left button press results in a binding between the color feature green and the action features of the left button press. If the consecutive trial then requires the participant to respond with a left button press to a red word, the stimulus-response binding of the previous trial is retrieved, since one of the features (left button press) is the same as in the previous trial. This leads to conflict, since the actual binding (red - left button press) does not match the retrieved binding (green - left button press). Due to such binding mechanisms, participants respond more slowly to partial repetitions than to repetitions or changes of the complete stimulus (Hommel, 2004;

Kahneman et al., 1992; but see Rothermund, Wentura, & De Houwer, 2005 for an alternative account). These and likewise intertrial effects have been demonstrated for relevant and irrelevant features of visual stimuli (e.g., Dreisbach & Haider, 2008; Hommel & Keizer, 2012; Hommel, 1998; Kahneman et al., 1992; Kleinsorge, 1999; Zehetleitner, Rangelov, & Müller, 2012) and nonvocal auditory stimuli (Frings, Schneider, & Moeller, 2014; Mayr & Buchner, 2006; Moeller, Rothermund, & Frings, 2012; Zmigrod & Hommel, 2009, 2010). Although there is already some evidence for the integration of speaker identity in action-effect bindings (Herwig & Waszak, 2012) and in episodic events of a conflict task (Spapé & Hommel, 2008), a systematic investigation of stimulus-response bindings with regard to voice features (e.g., linguistic content, speaker identity, affective information) and action features of spoken words is still lacking. Especially the stimulus-response integration of vocal affective information that is transferred through emotional prosody has not been addressed before. Therefore, as a first aim, the present study investigates whether binding processes reassemble the voice features and action features of spoken words (as indicated by partial repetition costs) with regard to speaker identity (Experiment 1), and affective prosody (Experiment 2) of spoken word stimuli.

Moreover, we were interested in the flexibility and boundary conditions of voice-related binding processes. Previous research has shown that partial repetition costs as a consequence of binding processes between visual stimulus features and response features are modulated by the kind of task representation (stimulus-response mappings vs. general task rule) participants adopt. In a study by Dreisbach and Haider (2008), two groups of participants had to respond to an identical stimulus set of eight words with two responses (four stimuli were mapped to a left and four stimuli were mapped to a right response). The only difference was the task representation of the groups. One group of participants were instructed to respond to the eight words according to predefined stimulus-response rules (SR group) with either a left or a right button press (e.g., “bug” - left button, “sofa” - right button). The other group of participants were instructed to respond to the eight words according to a simple task rule (TR group) that defines a response discriminating stimulus feature (“whenever the object is a moving object press the left button, if it is a non-moving object, press the right button”) and thus allows the grouping of the stimulus set into those stimuli that afford a left response and those that afford a right response. Irrelevant for the

assigned response, stimulus words were presented either in red or green. In the SR group sequential intertrial analyses revealed typical partial repetition costs: participants responded more slowly to partial stimulus repetitions (response repetition along with color shift or response shift along with color repetition) than to repetitions or changes of the complete stimulus (response repetition along with color repetition or response shift along with color shift). This means, in the SR group binding occurred between the irrelevant color feature and the response feature of the stimuli. Importantly, in the TR group, same analyses did not reveal partial repetition costs, indicating that the application of a task rule that defines the response discriminating stimulus feature (e.g., moving vs. non-moving) modulated binding processes between any other stimulus features and the response. This effect has been termed the *shielding function* of task rules (Dreisbach & Haider, 2008, 2009; Dreisbach, 2012). So far, however, the shielding function of task rules has only been demonstrated with neutral visual material. It still has to be shown whether the shielding mechanism also holds true for social-affective stimuli like spoken words. This cannot be taken for granted since spoken words are different from written words in many ways. The most important difference is the social-affective value of the human voice (Belin et al., 2011; Sumner, 2015). Information about speaker identity and the motivational and emotional state of the speaker is taken into account very early during language comprehension and influences the processing of linguistic content (Sauter & Eimer, 2010; Van Berkum, van den Brink, Tesink, Kos, & Hagoort, 2008). Furthermore, affective information in general appears to be harder to ignore than neutral information (e.g., Compton, 2003; Öhman, Flykt, & Esteves, 2001; but see Colzato, van Wouwe, & Hommel, 2007; Holmes, Kiss, & Eimer, 2006). If binding processes can be found between irrelevant voice features (i.e., speaker identity, emotional prosody) and response features, it needs to be investigated whether the application of task rules modulates these voice-related binding processes.

The present study was designed (1) to provide direct evidence for binding processes between neutral and affective voice features and action-related features of spoken word stimuli and (2) to investigate the flexibility of these voice-related binding processes (i.e. their modulation by the task representation). Therefore, we used eight stimulus words, four of which were mapped to a left and a right response, respectively. As irrelevant voice features we manipulated speaker identity by recording the eight stimulus words both with a male

speaker and with a female speaker (Experiment 1) and we manipulated affective information by recording the eight stimulus words both with an angry and a neutral prosody (Experiment 2). In each experiment, all participants had to respond to the very same vocal stimulus set with the only difference, that half of them were instructed to learn the respective stimulus-response rules by rote (SR group) and the other half were instructed to apply a specific task rule (e.g., “whenever the word represents something that fits/does not fit into a bag, press the left/right button”, TR group). If binding occurs for voice features and action-related features of spoken stimuli, we should find partial repetition costs in the SR group. In the TR group, the application of the task rule that defines the response relevant stimulus feature should modulate these binding processes. In sum, we therefore expect a voice (repetition vs. shift) by response (repetition vs. shift) interaction in the SR condition but not in the TR condition.

Experiment 1

Material and Methods

Participants

Forty participants (age $M = 23.23$ years, $SD = 4.07$ years, range 18–41 years, 32 females) from the University of Regensburg, Germany, participated for partial course credit or a small financial reward. Twenty participants were assigned randomly to each of the two experimental conditions. Participants signed informed consent and were debriefed after the session.

Stimuli and Procedure

The recordings of eight German one syllable words (book, house, shirt, knee, horse, mushroom, cord, time) served as stimuli. Each word was recorded twice, with a female and a male speaker. For half of the participants, the words *book*, *shirt*, *mushroom* and *cord* were assigned to a left response key (Y key on a QWERTZ keyboard) and the words *house*, *knee*, *horse* and *time* were assigned to a right response key (M key). For the second half of participants the assignment was reversed.

A given trial started with a fixation cross of 300 ms duration. Then, the target word was presented via headphones accompanied by a blank screen until the response was given. After an inter-trial interval (ITI) of 600 ms, the next trial started. When participants responded using the wrong button, they received visual feedback for 1500 ms.

The experiment consisted of 4 practice blocks and 3 experimental blocks. In the practice blocks, the stimuli were introduced stepwise (two words per block). In the first block, only two words were presented and in the following practice blocks, stimulus set size increased by two new stimuli, such that in the last practice block (block 4) all stimuli were presented. In the experimental blocks, as in block 4, all eight words were presented. Each word was presented ten times per block, five times spoken by the female voice and five times spoken by the male voice, resulting in a block length of 20 trials for the first block, 40 trials for the second block, 60 trials for the third block and 80 trials for Blocks 4-7. For each block, the list of target words was randomized, with the exception that it was ensured that there were no direct word repetitions and that there was an equal number of each factor combination (Voice x Response). At the beginning of each block, participants were informed about the two new words and the correct key assignments. After Block 4, they were told that no more words would appear and received a scheme that listed all eight words together with the correct key assignments. The general procedure was identical in the SR condition and the TR condition except for the instructions at the beginning. In the SR condition, participants were instructed that the experimenters were interested in the assignment of specific words to specific responses and the words were introduced with the assigned response keys. In the TR condition, participants were instructed that the experimenters were interested in the assignment of specific words to specific categories and were instructed to press the right/left key when the word presented something that fits into a bag and to press the other key whenever the word presented an object not fitting into a bag. This rather arbitrary task rule was chosen to prevent participants in the SR group from guessing the underlying task rule (see also Dreisbach, Goschke, & Haider, 2006, 2007). After the last experimental block, participants were asked to fill out a questionnaire. Participants in the SR condition were asked, whether they had used any recall tricks to memorize the eight words. Participants in the TR condition were asked whether they found the task rule useful and, if not, whether they had used any other strategy. Data of one participant in the SR group had

to be replaced by the data of an additional participant because of having created and used a task-rule¹. The experimental session lasted about 25 minutes.

Design

We applied a 2 (Voice_{identity}: repetition vs. shift) x 2 (Response: repetition vs. shift) x 2 (Instruction: SR vs. TR) mixed factors design. Instruction condition was implemented as a between factor; all other factors were implemented as within factors. Reaction times (RTs) and error rates served as dependent variables.

Results and Discussion

Preprocessing

Erroneous and post error trials (8.9 %) as well as the first trial of each block were excluded from RT analyses. Furthermore, all RTs more than three standard deviations different from the individual cell mean (1.8 %) were excluded. Data were collapsed across the three experimental blocks (Block 5-7).

RT data

Figure 1 plots mean RTs as a function of Instruction, Voice_{identity} and Response. We conducted a 2 (Instruction: SR vs. TR) x 2 (Response: repetition vs. shift) x 2 (Voice_{identity}: repetition vs. shift) mixed factors ANOVA with repeated measures for the latter two factors. This revealed a significant main effect of Response, $F(1,38) = 10.41, p = .003, \eta_p^2 = .22$, indicating faster responses when the response shifted compared to when the response was repeated (841 ms vs. 862 ms). Furthermore, there was a significant interaction Response x Voice_{identity}, $F(1,38) = 15.92, p < .001, \eta_p^2 = .30$, and a marginal significant interaction Instruction x Response, $F(1,38) = 3.23, p = .081, \eta_p^2 = .08$. Most importantly with respect to our hypotheses, the triple interaction Instruction x Voice_{identity} x Response was significant, $F(1,38) = 6.64, p = .014, \eta_p^2 = .15$. None of the other effects was significant (all $F_s < 0.9$).

¹ In a previous study (Dreisbach & Haider, 2009) we could show that participants who receive SR instructions but create their own task rule actually behave like participants in the TR condition. Since here, we do not know at what point in time during the experiment the participant created and used his/her own task rule, we decided to exclude these data entirely and replace it with data of an additional participant.

Separate 2x2 ANOVAs for each instruction group revealed that the significant triple interaction can be described by a significant Voice_{identity} x Response interaction in the SR group, $F(1,19) = 19.7, p < .001, \eta_p^2 = .51$, and a non-significant interaction in the TR group $F(1,19) = 1.10, p = .306, \eta_p^2 = .06$. This indicates that only in the SR group, at response repetitions, performance was quicker if voice identity also repeated than if voice identity shifted, and at response shifts, performance was quicker if voice identity also shifted than if it was repeated, whereas such interaction was absent in the TR group (Figure 1).

Error data

Mean percentages of errors are plotted in Figure 1 as a function of Instruction group, Voice_{identity} and Response. A 2 (Instruction: SR vs. TR) x 2 (Response: repetition vs. shift) x 2 (Voice_{identity}: repetition vs. shift) mixed factors ANOVA with repeated measures for the latter two factors yielded a main effect of Response, $F(1,38) = 8.52, p = .006, \eta_p^2 = .18$, indicating that, overall, response repetitions were more error prone than response shifts (5.3 % vs. 3.7 %). Furthermore, a significant main effect for Instruction, $F(1,38) = 12.58, p = .001, \eta_p^2 = .25$, revealed that participants in the TR group committed fewer errors than participants in the SR group (2.2 % vs. 6.8 %). The interaction Response x Voice_{identity} also reached significance, $F(1,38) = 9.98, p = .003, \eta_p^2 = .21$, and the interaction Instruction x Response reached marginal significance $F(1,38) = 3.86, p = .057, \eta_p^2 = .09$. Again, there was a significant triple interaction Instruction x Voice_{identity} x Response, $F(1,38) = 8.08, p = .007, \eta_p^2 = .18$, indicating a significant interaction Response x Voice_{identity} in the SR group, $F(1,19) = 11.7, p = .003, \eta_p^2 = .38$, and a non-significant interaction Response x Voice_{identity} in the TR group, $F(1,19) = .11, p = .747, \eta_p^2 = .006$. No further effect was significant (all F s < 1.26).

– Insert Figure 1 about here –

The results of Experiment 1 clearly support our assumptions. The triple interactions in RT and error data confirmed the significant interaction of Voice_{identity} by Response in the SR condition and the lack thereof in the TR condition. This suggests that binding processes reassemble voice features and action-related features of spoken stimuli (SR condition).

Moreover, the application of a task rule modulated these binding effects (TR condition). Note, that overall error rates were generally very low ($< 5\%$) and should therefore be interpreted with caution.².

Experiment 1 demonstrated binding between speaker identity and action-features of spoken stimuli and its modulation by task representations. According to the model of voice perception of Belin and colleagues (2004), there is another important feature of vocal stimuli that is voice-specific and is processed in specialized areas: vocal affective information. In order to investigate if binding between voice features and action features and its modulation by task representations also occurs for vocal affective information, we conducted a second experiment. We recorded eight neutral words twice, with the same male speaker but with two different affective prosodies (neutral and angry). As in Experiment 1, irrelevant of the voice feature, participants had to respond to these spoken word stimuli according to predefined stimulus-response rules (SR group) or according to a task rule (TR group). That way, for the first time, we could investigate binding processes between vocal affective information and action-related features of spoken words and test whether the shielding function of task rules would also hold for social-affective unrelated information.

² In order to test how reliable the effect is, we replicated the findings of Experiment 1 in an additional experiment with stimuli recorded by a pre-pubertal female and a pre-pubertal male child. We tested 40 participants (20 in the SR group, 20 in the TR group). The procedure was exactly the same as in Experiment 1 with the exception that the stimuli were recorded by a pre-pubertal female and a pre-pubertal male child. As in Experiment 1, RT analyses revealed a significant triple interaction Instruction $\times$ Voice_{identity} $\times$ Response, $F(1,38) = 5.42$, $p = .025$, $\eta_p^2 = .13$. Separate 2x2 ANOVAs for each Instruction group revealed a significant Voice_{identity} $\times$ Response interaction in the SR group, $F(1,19) = 9.96$, $p = .005$, $\eta_p^2 = .34$, and a non-significant interaction in the TR group $F(1,19) = .56$, $p = .463$, $\eta_p^2 = .03$. In the Error data, the triple interaction Instruction $\times$ Voice_{identity} $\times$ Response was not significant, $F(1,38) = .19$, $p = .669$, $\eta_p^2 = .01$. Overall error rates were generally very low ($< 3\%$) and should be interpreted with caution.

Experiment 2

Material and Methods

Participants

Forty participants (age $M = 22.08$ years, $SD = 2.74$ years, range 18–30 years, all females) from the University of Regensburg, Germany, participated for partial course credit or a small financial reward. Twenty participants were assigned to each of the two instruction conditions. Participants signed informed consent and were debriefed after the session. None of the participants had participated in Experiment 1.

Stimuli and procedure

The recordings of eight German two syllable words (stomach, shelf, bucket, list, pea, oar, thumb, flag) served as stimuli. All words were taken from the Berlin affective word-list BAWL-R (Vö et al., 2009) and had a neutral mean valence rating of 0.26 (range: -0.35 - 0.60) on a 7-point scale ranging from -3 (very negative) through 0 (neutral) to +3 (very positive) and a low mean arousal rating of 2.04 (range: 1.75 - 2.44) on a 5-point scale ranging from 1 (low arousal) to 5 (high arousal). Each word was recorded twice with a male speaker and two different affective prosodies (neutral and angry). The recordings were normalized to ensure comparable loudness of neutral and angry stimuli. The valence and arousal of each stimuli's prosody were rated by all participants after the experimental procedure using a 9-point version of the Self-Assessment Manikin Scale (SAM; Bradley & Lang, 1994). This revealed a mean valence rating of -0.25 for the stimuli spoken with a neutral prosody and a mean valence rating of -1.75 for the stimuli spoken with an angry prosody on a 9-point scale ranging from -4 (very negative) through 0 (neutral) to +4 (very positive). A 2 (Prosody: neutral vs. angry) x 2 (Instruction: SR vs. TR) ANOVA with the valence rating as dependent variable revealed a main effect Prosody, $F(1,38) = 70.48$, $p < .001$, $\eta_p^2 = .65$. The main effect Instruction and the interaction of the two factors did not reach significance (all F s < 1.50). The arousal rating revealed a mean arousal rating of 1.72 for the stimuli spoken with a neutral prosody and a mean arousal rating of 6.82 for the stimuli spoken with an angry prosody on a 9-point scale ranging from 1 (low arousal) to 9 (high arousal). A 2 (Prosody: neutral vs. angry) x 2 (Instruction: SR vs. TR) ANOVA with the arousal rating as dependent

variable revealed a main effect Prosody, $F(1,38) = 1180.46$, $p < .001$, $\eta_p^2 = .97$. The main effect Instruction and the interaction of the two factors did not reach significance (all F s $< .86$).

The experimental procedure was exactly the same as in Experiment 1, except the task rule in the TR condition: Participants in the TR group were instructed to press the right/left key when the word presented something that can be filled with content and to press the other key whenever the word presented something that cannot be filled with content. None of the participants in the SR condition guessed the underlying task rule or generated an own rule. Participants in the TR group all found the rule useful. Thus all participants were included in the analysis.

Design

A 2 (Voice_{affect}: repetition vs. shift) $\times$ 2 (Response: repetition vs. shift) $\times$ 2 (Instruction condition: SR vs. TR) mixed factors design was applied with Response and Voice_{affect} as the within-subject-factors and Instruction as the between-subject-factor. RTs and error rates served as dependent measures.

Results and Discussion

Preprocessing

Erroneous and post error trials (11.3 %) and the first trial of each block were excluded from RT analyses. Furthermore, all RTs differing more than three standard deviations from the individual cell mean (1.8 %) were excluded.

RT data

Figure 2 plots mean RTs as a function of Instruction group, Voice_{affect} and Response. We conducted a 2 (Instruction: SR vs. TR) $\times$ 2 (Response: repetition vs. shift) $\times$ 2 (Voice_{affect}: repetition vs. shift) mixed factors ANOVA with repeated measures on the latter two factors. We found a significant main effect of Response, $F(1,38) = 6.37$, $p = .016$, $\eta_p^2 = .14$, indicating faster responses for response shifts than for response repetitions (856 ms vs. 876 ms) and a significant main effect of Voice_{affect}, $F(1,38) = 6.45$, $p = .015$, $\eta_p^2 = .15$, indicating faster responses for Voice_{affect} repetitions than for Voice_{affect} shifts (858 ms vs. 875 ms).

Furthermore, a significant main effect for Instruction, $F(1,38) = 4.49, p = .041, \eta_p^2 = .11$, revealed that participants of the TR group responded faster than participants in the SR group (811 ms vs. 921 ms). The interaction Response x Voice_{affect} was also significant, $F(1,38) = 4.75, p = .036, \eta_p^2 = .11$. Most importantly, the triple interaction Instruction x Voice_{affect} x Response was also significant, $F(1,38) = 8.25, p = .007, \eta_p^2 = .18$. None of the other effects was significant (all $F_s < 1.02$). Separate 2x2 ANOVAs for each Instruction group revealed a significant Voice_{affect} x Response interaction in the SR group, $F(1,19) = 9.60, p = .006, \eta_p^2 = .34$, and a non-significant interaction in the TR group $F(1,19) = .36, p = .557, \eta_p^2 = .02$. This indicates that in the SR group, response repetitions facilitated performance if voice affect also repeated and impaired performance if voice affect shifted, whereas in the TR group, such interaction was absent.

Error data

Mean percentages of errors are plotted in Figure 2 as a function of Instruction group, Voice_{affect} and Response. A 2 (Instruction: SR vs. TR) x 2 (Response: repetition vs. shift) x 2 (Voice_{affect}: repetition vs. shift) mixed factors ANOVA with repeated measures on the last two factors yielded a main effect of Response, $F(1,38) = 15.19, p < .001, \eta_p^2 = .29$, indicating that, overall, response repetitions were more error prone than response shifts (6.7 % vs. 4.6 %). Furthermore, a significant main effect for Instruction, $F(1,38) = 9.42, p = .004, \eta_p^2 = .20$, revealed that participants in the TR group committed fewer errors than participants in the SR group (2.7 % vs. 8.5 %). The interaction Response x Voice_{affect} also reached significance, $F(1,38) = 11.97, p = .001, \eta_p^2 = .24$. The triple interaction Instruction x Voice_{affect} x Response was not significant, $F(1,38) = 2.44, p = .127, \eta_p^2 = .06$. No further effect was significant (all $F_s < .60$).

– Insert Figure 2 about here –

RT data from Experiment 2 replicated our findings from Experiment 1 with affective prosody as irrelevant voice feature. Again, we found a significant interaction of the irrelevant voice feature and the response in the SR condition, indicating binding between voice

features and response-related features of spoken stimuli. And, as in Experiment 1, the triple interaction with the instruction condition showed that the application of a task rule reduced the impact of the irrelevant voice features: binding effects were again virtually absent. Results of error data did not reveal any relevant significant effects. They were generally low (< 6%) and did not contradict any of the RT results.

General Discussion

The experiments of the present study successfully demonstrated binding processes between voice features and response features of spoken word stimuli. Furthermore, results support our assumption that these binding processes are modulated by the specific task representation participants adopt.

With regard to the SR groups, our experiments provide clear evidence for binding processes between voice and response features. In both experiments, analyses of the SR condition revealed typical partial repetition costs, indicated by an interaction between the irrelevant voice feature (Experiment 1: speaker identity; Experiment 2: vocal affective information) and the response: performance was worse if only one feature – voice feature or response – repeated and the other feature shifted and was better when both features shifted or both repeated. These effects can be explained in terms of feature binding: Responding to the spoken word created a transient binding between the voice feature (male or female speaker identity in Experiment 1; negative or neutral affective prosody in Experiment 2) and the response (left or right). This binding then got retrieved when one or both features of the prior episode repeated and consequently incurred costs for partial repetitions and benefits for full repetitions. The results of Experiment 1 offer the first direct evidence for the integration of speaker identity in stimulus-response bindings. Therefore it complements previous research that hint at the integration of speaker identity in action effect-bindings (Herwig & Waszak, 2012) and in episodic events of a conflict task (Spapé & Hommel, 2008). Moreover the results yielded by Experiment 2 denote for the first time the integration of vocal affective information in stimulus-response events. It is important to note, that in all experiments the voice feature was completely irrelevant for the task at hand. Nevertheless, as indicated by our results, it was bound to the response that has been

made according to the stimulus-response rules. This is in line with studies demonstrating that feature-response bindings also involve irrelevant visual stimulus features (Dreisbach & Haider, 2008; Hommel, 2005), irrelevant nonvocal auditory stimulus features (e.g., Frings et al., 2014; Mayr & Buchner, 2006; Moeller et al., 2012; Zmigrod & Hommel, 2009, 2010) or visual distractors (Frings, Rothermund, & Wentura, 2007; Frings & Rothermund, 2011; Giesen, Frings, & Rothermund, 2012).

With regard to the TR groups of our experiments, the data support the assumption that binding processes between voice features and response-related features of spoken stimuli are modulated by the application of a task rule that defines the response-discriminating stimulus feature of a given stimulus-set. In contrast to the SR groups, the interaction between voice and response features was virtually absent in the TR groups in both experiments. This result can be interpreted as evidence for task rules generally impairing the creation of a binding between irrelevant stimulus features (here: speaker identity and vocal affective information) and the response. Alternatively, the creation of bindings could be unaffected, but the task rule might impair the retrieval of the potentially interfering speaker-response binding created in the previous trial. With regard to previous studies on the application of task rules, we assume that the impairment of the binding process is the more likely alternative. A task rule defines a response-discriminating stimulus feature of a given stimulus set. It has been shown previously, that the application of a task rule promotes prioritized processing of task related (i.e. response defining) stimulus information which might go along with reduced processing of irrelevant stimulus information not being part of the task rule (Reisenauer & Dreisbach, 2013, 2014). In terms of theories involving attentional feature weighting, in our experiments, the application of the task rule might have increased the attentional weights of semantic feature codes and at the same time decreased the attentional weights of task-irrelevant feature codes (cf. Bundesen, Habekost, & Kyllingsbæk, 2005; Bundesen, 1990; Desimone & Duncan, 1995; Hommel, Müsseler, Aschersleben, & Prinz, 2001; Müller, Heller, & Ziegler, 1995). As a consequence the code activation of speaker identity might have become too low to pass a kind of an integration threshold and therefore might have missed getting bound to other activated feature or response codes (cf. Hommel, 2004). Alternatively, it has been suggested that a temporal overlap of code activation is mandatory for the creation of a binding and that

attentional weights do not only influence the peak but also the duration of code activation (Zmigrod & Hommel, 2009). Speaker identity and vocal affective information have been shown to be encoded at a very early stage of speech processing (Sauter & Eimer, 2010; Van Berkum et al., 2008). In the SR condition, the duration of code activation might have been long enough to overlap with the later activation of the response code. But in the TR condition, with decreased attentional weights, the early code activations of speaker identity and vocal affective information might have been too short to overlap with the response code and therefore, the binding of speaker identity to the response code might have been prevented.

In sum, the experiments presented in this study extend previous research by demonstrating that binding mechanisms that have been found to reassemble relevant and irrelevant stimulus features and action-related features of visual events (e.g., Dreisbach & Haider, 2008; Frings et al., 2007; Frings & Rothermund, 2011; Giesen et al., 2012; Hommel, 1998; Kahneman et al., 1992; Kleinsorge, 1999; Zehetleitner et al., 2012) and nonvocal auditory events (e.g., Frings et al., 2014; Mayr & Buchner, 2006; Moeller et al., 2012; Zmigrod & Hommel, 2009, 2010) also reassemble voice features and action-related features of vocal events, the most complex multi-feature events of our everyday life. Moreover, our results indicate that these voice-related binding processes are modulated by the adopted task representation and therefore demonstrate that the *shielding function* of task rules can be generalized from neutral visual events (Dreisbach & Haider, 2008, 2009; Dreisbach, 2012) to the processing of auditory social-affective events like spoken words. On a more general level, the results of this study are suggestive of a very adaptive binding mechanism that ensures the coherence and flexibility of speech perception: without a specific rule (SR condition), binding between voice and action-related features enables a coherent perception of multiple social-affective information that is offered by vocal events. The absence of binding (or the retrieval thereof) when having a specific rule in mind on the other side enables extracting task rule related information irrespective of irrelevant vocal information. Depending on the current task context, both mechanisms are adaptive and warrant successful communication and action in a dynamic social environment.

Acknowledgements

We would like to thank Lisa Raith, Katharina Kellner, Katharina Konopka and Johanna Ganslmeier for data collection and the speakers for recording the stimuli. This research was funded by a grant to the last author from the Deutsche Forschungsgemeinschaft (DFG; project DR 392/5-3).

Compliance with Ethical Standards

The authors declare that they have no conflict of interest. All procedures performed in this study were in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study.

References

- Belin, P., Bestelmeyer, P. E. G., Latinus, M., & Watson, R. (2011). Understanding voice perception. *British Journal of Psychology*, 102(4), 711–25.
- Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends in Cognitive Sciences*, 8(3), 129–35.
- Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, 403(6767), 309–312.
- Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1), 49–59.
- Bundesen, C. (1990). A theory of visual attention. *Psychological Review*, 97(4), 523–547.
- Bundesen, C., Habekost, T., & Kyllingsbæk, S. (2005). A neural theory of visual attention: bridging cognition and neurophysiology. *Psychological Review*, 112(2), 291–328.
- Colzato, L. S., van Wouwe, N. C., & Hommel, B. (2007). Feature binding and affect: emotional modulation of visuo-motor integration. *Neuropsychologia*, 45(2), 440–6.
- Compton, R. J. (2003). The Interface Between Emotion and Attention: A Review of Evidence From Psychology and Neuroscience. *Behavioral and Cognitive Neuroscience Reviews*, 2(2), 115–129.
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18, 193–222.
- Dreisbach, G. (2012). Mechanisms of cognitive control: the functional role of task rules. *Current Directions in Psychological Science*, 21(4), 227–231.
- Dreisbach, G., Goschke, T., & Haider, H. (2006). Implicit task sets in task switching? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(6), 1221–33.
- Dreisbach, G., Goschke, T., & Haider, H. (2007). The role of task rules and stimulus-response mappings in the task switching paradigm. *Psychological Research*, 71(4), 383–92.
- Dreisbach, G., & Haider, H. (2008). That's what task sets are for: shielding against irrelevant

- information. *Psychological Research*, 72(4), 355–61.
- Dreisbach, G., & Haider, H. (2009). How task representations guide attention: further evidence for the shielding function of task sets. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(2), 477–86.
- Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1(1), 1–47.
- Formisano, E., Martino, F. De, Bonte, M., & Goebel, R. (2008). “Who” is saying “what”? Brain-based decoding of human voice and speech. *Science*, 322, 970–973.
- Frings, C., & Rothermund, K. (2011). To be or not to be...included in an event file: integration and retrieval of distractors in stimulus-response episodes is influenced by perceptual grouping. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(5), 1209–27.
- Frings, C., Rothermund, K., & Wentura, D. (2007). Distractor repetitions retrieve previous responses to targets. *Quarterly Journal of Experimental Psychology*, 60(10), 1367–77.
- Frings, C., Schneider, K. K., & Moeller, B. (2014). Auditory distractor processing in sequential selection tasks. *Psychological Research*, 78, 411–422.
- Giesen, C., Frings, C., & Rothermund, K. (2012). Differences in the strength of distractor inhibition do not affect distractor-response bindings. *Memory & Cognition*, 40(3), 373–87.
- Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., & Vuilleumier, P. (2005). The voices of wrath: brain responses to angry prosody in meaningless speech. *Nature Neuroscience*, 8(2), 145–6.
- Heilman, K. M., Scholes, R., & Watson, R. T. (1975). Auditory affective agnosia. Disturbed comprehension of affective speech. *Journal of neurology, neurosurgery, and psychiatry*, 38(1), 69–72.
- Herwig, A., & Waszak, F. (2012). Action-effect bindings and ideomotor learning in intention and stimulus-based actions. *Frontiers in Psychology*, 3:444.
- Holmes, A., Kiss, M., & Eimer, M. (2006). Attention modulates the processing of emotional

- expression triggered by foveal faces. *Neuroscience Letters*, 394(1), 48–52.
- Hommel, B. (1998). Event Files: evidence for automatic integration of stimulus-response episodes. *Visual Cognition*, 5(1-2), 183–216.
- Hommel, B. (2004). Event files: feature binding in and across perception and action. *Trends in Cognitive Sciences*, 8(11), 494–500.
- Hommel, B. (2005). How much attention does an event file need? *Journal of Experimental Psychology: Human Perception and Performance*, 31(5), 1067–82.
- Hommel, B., & Keizer, A. W. (2012). Binding success and failure: Evidence for the spontaneous integration of perceptual features and object evaluations. *Frontiers in Psychology*, 3:581.
- Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The Theory of Event Coding (TEC): a framework for perception and action planning. *Behavioral and Brain Sciences*, 24(5), 849–78; discussion 878–937.
- Jeannerod, M. (1999). To act or not to act: perspectives on the representation of actions. *Quarterly Journal of Experimental Psychology*, 52(1), 1–29.
- Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: object-specific integration of information. *Cognitive Psychology*, 24(2), 175–219.
- Kleinsorge, T. (1999). Response repetition benefits and costs. *Acta Psychologica*, 103, 295–310.
- Kriegstein, K. V., & Giraud, A. L. (2004). Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *NeuroImage*, 22(2), 948–955.
- Mayr, S., & Buchner, A. (2006). Evidence for episodic retrieval of inadequate prime responses in auditory negative priming. *Journal of Experimental Psychology: Human Perception and Performance*, 32(4), 932–43.
- Moeller, B., Rothermund, K., & Frings, C. (2012). Integrating the irrelevant sound. *Experimental Psychology*, 59(5), 258–264.
- Müller, H. J., Heller, D., & Ziegler, J. (1995). Visual search for singleton feature targets within and across feature dimensions. *Perception & psychophysics*, 57(1), 1–17.

- Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention_Snakes in the grass.pdf. *Journal of Experimental Psychology: General*, 130(3), 466–478.
- Reisenauer, R., & Dreisbach, G. (2013). The impact of task rules on distracter processing: automatic categorization of irrelevant stimuli. *Psychological Research*, 77, 128–138.
- Reisenauer, R., & Dreisbach, G. (2014). The shielding function of task rules in the context of task switching. *Quarterly Journal of Experimental Psychology*, 67(2), 358–376.
- Rothermund, K., Wentura, D., & De Houwer, J. (2005). Retrieval of incidental stimulus-response associations as a source of negative priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(3), 482–495.
- Sauter, D. A., & Eimer, M. (2010). Rapid detection of emotion from human vocalizations. *Journal of Cognitive Neuroscience*, 22(3), 474–81.
- Schirmer, A., Zysset, S., Kotz, S. a., & Yves von Cramon, D. (2004). Gender differences in the activation of inferior frontal cortex during emotional speech perception. *NeuroImage*, 21(3), 1114–23.
- Spapé, M. M., & Hommel, B. (2008). He said, she said: episodic retrieval induces conflict adaptation in an auditory Stroop task. *Psychonomic Bulletin & Review*, 15(6), 1117–21.
- Sumner, M. (2015). The social weight of spoken words. *Trends in Cognitive Sciences*, 19(5), 238–239.
- Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology*, 6(2), 171–8.
- Tryon, W. W. (2001). Evaluating statistical difference, equivalence, and indeterminacy using inferential confidence intervals: an integrated alternative method of conducting null hypothesis statistical tests. *Psychological Methods*, 6(4), 371–86.
- Van Berkum, J. J. a, van den Brink, D., Tesink, C. M. J. Y., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, 20(4), 580–91.
- Van Lancker, D. R., & Canter, G. J. (1982). Impairment of voice and face recognition in patients with hemispheric damage. *Brain and Cognition*, 1(2), 185–95.
- Van Lancker, D. R., Kreiman, J., & Cummings, J. (1989). Voice perception deficits:

neuroanatomical correlates of phonagnosia. *Journal of clinical and experimental neuropsychology*, 11(5), 665–674.

Võ, M. L.-H., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., & Jacobs, A. M. (2009). The Berlin Affective Word List Reloaded (BAWL-R). *Behavior Research Methods*, 41(2), 534–8.

Zehetleitner, M., Rangelov, D., & Müller, H. J. (2012). Partial repetition costs persist in nonsearch compound tasks: evidence for multiple-weighting-systems hypothesis. *Attention, Perception, & Psychophysics*, 74(5), 879–90.

Zmigrod, S., & Hommel, B. (2009). Auditory event files: integrating auditory perception and action planning. *Attention, Perception, & Psychophysics*, 71(2), 352–62.

Zmigrod, S., & Hommel, B. (2010). Temporal dynamics of unimodal and multimodal feature binding. *Attention, Perception, & Psychophysics*, 72(1), 142–152.

Figure Captions

Figure 1. Response repetition effects averaged across the three experimental blocks for the SR group and the TR group in Experiment 1: mean RTs (ms) and error rates (%) are shown as a function of voice sequence (repetition vs. shift) and response sequence (repetition vs. shift). *Error bars* represent inferential confidence intervals (Tryon, 2001) based on the corresponding voice repetition vs. voice shift comparison.

Figure 2. Response repetition effects averaged across the three experimental blocks for the SR group and the TR group in Experiment 2: mean RTs (ms) and error rates (%) are shown as a function of voice sequence (repetition vs. shift) and response sequence (repetition vs. shift). *Error bars* represent inferential confidence intervals (Tryon, 2001) based on the corresponding voice repetition vs. voice shift comparison.