Atten Percept Psychophys (2016) 78:1488–1495
DOI 10.3758/s13414-016-1119-2

To hear or not to hear: Voice processing under visual load

Romi Zäske[1,2] · Marie-Christin Perlich[1] · Stefan R. Schweinberger[1]

Published online: 5 May 2016
© The Psychonomic Society, Inc. 2016

Abstract Adaptation to female voices causes subsequent voices to be perceived as more male, and vice versa. This contrastive aftereffect disappears under spatial inattention to adaptors, suggesting that voices are not encoded automatically. According to Lavie, Hirst, de Fockert, and Viding (2004), the processing of task-irrelevant stimuli during selective attention depends on perceptual resources and working memory. Possibly due to their social significance, faces may be an exceptional domain: That is, task-irrelevant faces can escape perceptual load effects. Here we tested voice processing, to study whether voice gender aftereffects (VGAEs) depend on low or high perceptual (Exp. 1) or working memory (Exp. 2) load in a relevant visual task. Participants adapted to irrelevant voices while either searching digit displays for a target (Exp. 1) or recognizing studied digits (Exp. 2). We found that the VGAE was unaffected by perceptual load, indicating that task-irrelevant voices, like faces, can also escape perceptual-load effects. Intriguingly, the VGAE was increased under high memory load. Therefore, visual working memory load, but not general perceptual load, determines the processing of task-irrelevant voices.

Keywords Voice · Gender · Aftereffects · Attention · Working memory · Perceptual load

Human voices are rich in social information about a speaker's identity, age, or gender (Schweinberger, Kawahara, Simpson, Skuk, & Zäske, 2014). Listeners routinely extract such cues even from nonspeech utterances (Skuk & Schweinberger, 2013b) or previously unheard speech (Zäske, Volberg, Kovács, & Schweinberger, 2014). Humans are often exposed to voices while engaging in other tasks, such as reading a newspaper in a busy coffee shop. The challenge for our attentional system is to focus on the task at hand, while monitoring the environment for behaviorally relevant information. The questions of whether and to what extent unattended voices are processed while we perform visual tasks is highly relevant for understanding both everyday voice perception and the distribution of attention between modalities.

Recent research on auditory adaptation suggested that exposure to nonlinguistic social cues in voices temporarily alters our perception of subsequent voices. For instance, prolonged listening to female voices causes androgynous test voices to sound more male, and vice versa (Schweinberger et al., 2008), suggesting contrastive coding of voice gender. Subsequent reports of voice aftereffects have revealed the neuronal codings of vocal age, identity, and affective information (Bestelmeyer, Rouger, DeBruine, & Belin, 2010; Skuk & Schweinberger, 2013a; Zäske, Schweinberger, & Kawahara, 2010; Zäske, Skuk, Kaufmann, & Schweinberger, 2013), in analogy to face aftereffects (reviewed in Webster & MacLeod, 2011). However, little is known about the role of attention in voice adaptation (but see Zäske, Fritz, & Schweinberger, 2013).

Adaptation has traditionally been conceived of as purely stimulus-driven. Accordingly, linguistic aftereffects were

Electronic supplementary material The online version of this article (doi:10.3758/s13414-016-1119-2) contains supplementary material, which is available to authorized users.

✉ Romi Zäske
romi.zaeske@uni-jena.de

[1] Department for General Psychology and Cognitive Neuroscience and DFG Research Unit Person Perception, Institute of Psychology, Friedrich Schiller University of Jena, Am Steiger 3, Haus 1, 07743 Jena, Germany

[2] Department of Otorhinolaryngology, Jena University Hospital, Lessingstraße 2, 07740 Jena, Germany

shown to be independent of focused attention to adaptors (Baart & Vroomen, 2010; Mullennix, 1986; Samuel & Kat, 1998; Sussman, 1993). At variance with these findings, the voice gender aftereffect (VGAE; Schweinberger et al., 2008) is abolished when spatial attention is diverted from adaptor voices (Zäske, Fritz, & Schweinberger, 2013). In Zäske, Fritz, and Schweinberger's study, participants simultaneously adapted to male or female voices in one ear and to gender-neutral (androgynous) voices in the other ear. They attended either the left or the right ear and classified voice gender (Exp. 1) or syllable (Exp. 2) of the adaptor voices. Irrespective of the task during adaptation, gender classifications of the subsequent test voices indicated a VGAE only when gender-specific (male or female) adaptors, but not when androgynous adaptors, had been spatially attended. Although this suggests that voice gender is not processed automatically during selective attention to another voice, it is unclear whether the VGAE is also modulated by selective attention to other stimuli, and to *visual* stimuli in particular.

Here we explored this question by manipulating visual selective attention during voice adaptation according to load theory (Lavie, Hirst, de Fockert, & Viding, 2004). This theory holds that the extent to which distractors are processed depends on both the availability of perceptual resources and working memory. Specifically, due to limits in attentional capacity, high *perceptual* load of a relevant task impairs distractor processing by leaving little capacity that automatically spills over to distractors. By contrast, high *working memory* load promotes distractor processing, by disrupting working memory control over target prioritization.

This account has received substantial support from studies on vision (reviewed in de Fockert, 2013; Lavie, 2005) and audition (Alain & Izenberg, 2003; Conway, Cowan, & Bunting, 2001; Dalton, Santangelo, & Spence, 2009; Fairnie, Moore, & Remington, 2016; Muller-Gass & Schröger, 2007; but see Murphy, Fraenkel, & Dalton, 2013), and from studies probing load theory for crossmodal attention (Berman & Colby, 2002; Brand-D'Abrescia & Lavie, 2008; Jacoby, Hall, & Mattingley, 2012; Macdonald & Lavie, 2011; Molloy, Griffiths, Chait, & Lavie, 2015; Raveh & Lavie, 2015; but see Tellinghuisen & Nowak, 2003). Interestingly, and at variance with load theory, several studies have suggested that faces present a special case, in the sense that they may recruit a domain-specific capacity-limited system (Neumann, Mohamed, & Schweinberger, 2011; Neumann & Schweinberger, 2008, 2009). Here we considered the possibility that voices are also "special" (Belin, Bestelmeyer, Latinus, & Watson, 2011) and might be relatively immune to perceptual load when unattended, similar to faces (Neumann & Schweinberger, 2008). At present, it is unclear whether a similar domain-specific attentional system exists for voices.

Of relevance for crossmodal situations, Moradi, Koch, and Shimojo (2005) showed that face processing is unaffected by auditory working memory load. Specifically, the magnitude of the face identity aftereffect was unaffected by the load of an auditory digit memory task in that study. Similarly, auditory aftereffects of adaptation to *linguistic* aspects of speech seem unaltered by visual task demands. For instance, Samuel and Kat (1998) reported that auditory aftereffects following adaptation to a phonetic [ba]–[wa] continuum were unaffected by visual attention to arithmetic or rhyming tasks, suggesting that speech adaptation is an automatic low-level process. Furthermore, Baart and Vroomen (2010) found aftereffects for a [b]–[d] continuum, irrespective of visuospatial or verbal working memory load during audiovisual adaptation. However, it is unclear whether *nonlinguistic* voice aftereffects would be susceptible to different visual task demands.

Here we tested whether irrelevant adaptor voices are processed despite visual selective attention to a perceptual (Exp. 1) or a working memory (Exp. 2) task. Previous findings suggested that spatial attention to androgynous voices abolishes the VGAE induced by unattended gender-specific voices (Zäske, Fritz, & Schweinberger, 2013). It is possible that an unattended voice is filtered in the presence of another attended voice, but would be processed in a standard perceptual-load task with alphanumeric character targets (similar to faces; Neumann & Schweinberger, 2008). Alternatively, and according to load theory, high perceptual load should leave relatively less attentional capacity to spill over to an ignored adaptor voice, thereby *impairing* its processing, and hence the VGAE (Lavie et al., 2004). Conversely, high working memory load should *increase* the VGAE, because it interferes with the maintenance of target prioritization. As a result, irrelevant adaptor voices should be increasingly processed.

Experiment 1

Method

Participants Thirty-two student participants (mean age = 21.9 years; range: 18–35; 16 female, four left-handed) contributed data. All of the participants were native German speakers, and none was familiar with any of the voices or reported hearing problems. All participants gave written informed consent and received course credit and an additional performance-based incentive of €1 or €2. The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of Friedrich Schiller University.

Stimuli The voice stimuli were audio recordings from five female and five male native German speakers (20–27 years of age) uttering the four vowel–consonant–vowel (VCV) syllables /aba/, /aga/, /ibi/, and /igi/. Voices were recorded by means of a Sennheiser MD 421-II microphone, a CEntrance MicPortPro preamplifier, and a SoundMax HD audio soundcard (16-bit resolution, sampling rate of 44.1 kHz). Recordings were normalized for average amplitude and adjusted to a uniform duration of 886 ms (including 100 ms silence at the beginning and end) using Adobe Audition 1.5 software.

These preprocessed voices were then set into five pairs of female and male voices and were entered into a morphing algorithm (Kawahara & Matsui, 2003). The pairings were matched for similarity in intensity patterns in the spectrogram, to increase morph quality. Four pairs were used for the experimental trials, and a fifth pair was used for practice trials only.

From each morphed pair, three stimuli were chosen as the androgynous test stimuli, corresponding to 40 %/60 %, 50 %/50 %, and 60 %/40 % female/male proportions. Thus, a total of 48 different test stimuli—that is, from each of the four VCV syllables and three morph levels (MLs) for each of the four female–male pairs—were available for the experimental trials. The two types of adaptor stimuli were VCV syllables spoken by the male (0 %/100 %) and female (100 %/0 %) voices from the same pairs as above.

Procedure Participants were tested individually in a dimly lit, sound-attenuated booth. Instructions and visual stimuli were delivered via a computer screen at a viewing distance of 65 cm. The visual stimuli were white digits presented on a black background. The digit arrays subtended a visual angle of 5.73° × 0.71°.

Voice stimuli were presented in mono via Sennheiser HD 212Pro headphones with an approximate peak intensity of 60 dB(A), as determined with a Brüel & Kjær Precision Sound Level Meter, Type 2206. The experimenter did not talk to the participants during the session, to avoid spurious adaptation effects. To keep participants motivated and focused on the selective attention task, they were told that they could receive an additional bonus of €1 or €2, contingent on their accuracy and speed in the visual task.

On each trial, participants performed a visual search task while hearing three irrelevant adaptor voices (Fig. 1).[1] Specifically, participants were asked to detect a 5 among an array of six digits (0–9). Depending on the adaptation block, concurrently presented task-irrelevant adaptor voices were either female or male. Following the offset of the third voice adaptor, participants classified an androgynous test voice according to gender. Female and male adaptation blocks were further subdivided into a low- and a high-load block, in which the degree of selective attention to the visual tasks was manipulated such that the six digits were either identical (low load) or all different (high load).

Trials started with a red fixation cross for 1,000 ms, followed by three identical voice adaptors for 886 ms each (including 100 ms of pre- and poststimulus silence). With the onset of the first adaptor, the fixation cross was replaced with a display of six horizontally arranged digits. Using the "d" and "l" keys of a computer keyboard (German layout), participants indicated as quickly and as accurately as possible whether a 5 was among the digits. There was a 60 % probability that a 5 was present. As soon as a response had been entered, a new display of digits appeared, and so on until the offset of the third voice adaptor. Thus, the number of test displays depended on the individual speed of participants. This was done to ensure constant attention to the digits. Following a black screen (300 ms) and a green question mark (2,000 ms), participants classified the test voice (886 ms) as quickly and as accurately as possible according to gender (female/male). Measured from voice onset, they had 2,886 ms to enter their response via the "d" and "l" keys before the question mark was replaced with a black screen (500 ms). If responses were too slow, the words "Please respond faster" appeared instead (500 ms). Thus, each trial lasted 9,444 ms.

The order of the adaptation blocks and load blocks in both experiments was counterbalanced across male and female participants. Morphed test voices (MLs 40 %/60 %, 50 %/50 %, and 60 %/40 %) were presented according to the method of constant stimuli. For a given trial, the adaptor and test voices always uttered VCV syllables that differed with respect to both vowels and consonants (e.g., /aba/ vs. /igi/). Also, the adaptor and test voices always originated from different speaker pairs. For instance, if a test voice was a morph from speaker pair #4, the preceding adaptor voices originated from speaker pairs #1, #2, or #3. This was done to ensure that any adaptation effects would indeed reflect high-level adaptation to voice quality, rather than low-level stimulus-dependent effects. There were 24 trials for each experimental condition (2 adaptation conditions × 2 load conditions). The nonexperimental factors Adaptor Syllable and Speaker Pair, as well as Test Syllable, Speaker Pair, and ML, were balanced such that all factor levels were equally often represented within each experimental block. After 24 trials, participants received a written feedback of their performance in the visual selective-attention task (i.e., number of correctly classified displays and mean reaction time [RT]).

Prior to the experiment the trial procedure was practiced stepwise with a fifth speaker pair not used in the main

[1] Note that in contrast to the present study, research on the linguistic aftereffects of speech has often used massed adaptation, with a relatively high number of adaptor stimuli followed by a complete series of test stimuli (e.g. Eimas & Corbit, 1973; Samuel & Kat, 1998). By contrast, we used only three adaptor stimuli preceding one test stimulus per trial. This was done in the tradition of previous studies on nonlinguistic adaptation, which indicated that a few adaptors are sufficient to elicit voice aftereffects for vocal gender, age, and identity (Schweinberger et al., 2008; Zäske, Schweinberger, & Kawahara, 2010; Zäske, Skuk, et al., 2013). Note also that our adaptor conditions were presented in blocks of 24 trials each, such that a much larger number (72 = 24 × 3) of adaptors of the same gender were interrupted only by the test stimuli. This design is more akin to one with "top-up" adaptors before each test stimulus, which is now also common in research on visual perceptual adaptation (e.g. Jenkins, Beaver, & Calder, 2006).

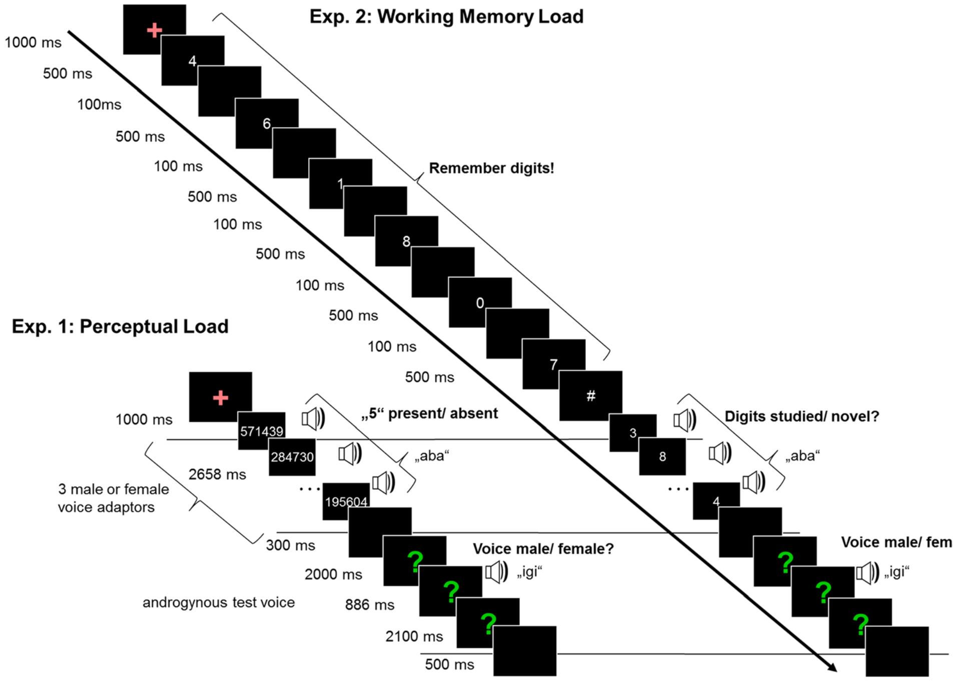

Fig. 1 Trial procedure in Experiments 1 and 2. Participants either performed a perceptual task (Exp. 1) or a working memory task (Exp. 2) on visually presented digits while adapting to task-irrelevant male or female voices. To manipulate selective attention during voice adaptation, the perceptual and working memory tasks were either easy (low load) or, as in the present example, more difficult (high load). Voice gender aftereffects, as induced by the ignored adaptor voices, were assessed by the perception of subsequent androgynous test voices as either male or female

experiment. In a first step (four trials), participants practiced the selective attention task without subsequent test voices. In a second step (ten trials), they were acquainted with the complete trial procedure. Overall, Experiment 1 lasted ~25 min.

Results

Validation of the load manipulation The successful manipulation of load was confirmed by analyses of variance (ANOVAs) with repeated measures on level of load (low/high) conducted for all performance measures in the selective attention task (Table 1): more correct displays under low than under high load [$F(1, 31) = 274.12$, $p < .001$, $\eta_p^2 = .898$], faster correct RTs during low than during high load [$F(1, 31) = 379.67$, $p < .001$, $\eta_p^2 = .925$], and more correct responses during low than during high load ($M = 94.3$ % vs. $M = 92.2$ %) [$F(1, 31) = 15.89$, $p < .001$, $\eta_p^2 = .339$]. Please note that accuracies were expected to be close to ceiling, due to the open response window, which allowed participants to search for the target digit at their own pace. The most informative measures of task difficulty are therefore the numbers of correct displays and correct RTs.

Table 1 Mean performance (*M*) and standard deviations (*SD*) in the selective attention task for low and high visual perceptual load, depicted separately for the mean number of displays (correct displays/total number), accuracy, and reaction times (RTs) for correctly classified displays

Level of Load	Displays (*n*)		Accuracy (%)		Correct RT (ms)	
	M	*SD*	*M*	*SD*	*M*	*SD*
Low	236/250	24	94.3	3.6	417	49
High	174/189	19	92.2	3.8	576	64

Voice gender aftereffects We performed an ANOVA on the proportions of “female” responses to androgynous test voices with repeated measures on adaptation condition (female/male) and level of load (low/high). Although ML effects were not the focus of the present study and were not analyzed for this reason, and due to the small number of trials, we provide a figure depicting values separately for each ML, which can be found in the supplemental information (Fig. S1). In short, adaptation effects appeared to be highly consistent across the tested MLs. We observed a significant VGAE, with more “female” responses following male than following female adaptation [$F(1, 31) = 62.93$, $p < .001$, $\eta_p^2 = .670$] ($M = 55.8$ % and 31.8 % “female” responses, respectively) and no effects of load (see Fig. 2). Please refer to Fig. 4 for the mean sizes of the aftereffects in Experiments 1 and 2.

Discussion

The VGAE was unaffected by the level of visual *perceptual* load, at variance with load theory (Lavie et al., 2004).

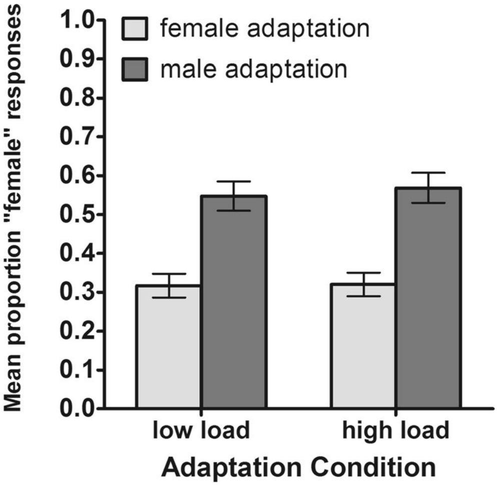

Fig. 2 Mean proportions of "female" responses following adaptation to female and male voices under low and high load in a visual search task (Exp. 1). Note that the voice aftereffects are not modulated by visual perceptual load. Error bars indicate *SEM*s

Accordingly, high relative to low perceptual load decreases the processing of task-irrelevant stimuli. For instance, inattentional deafness to simple tone stimuli can be induced by loading visual perceptual task demands (Macdonald & Lavie, 2011; Molloy et al., 2015; Raveh & Lavie, 2015). Accordingly, one might expect larger voice adaptation under low (vs. high) perceptual load in the present study, provided that attentional resources are shared by the target and distractor stimuli. However, resources may not always be shared between stimuli when the targets and distractors belong to different modalities (e.g., Allport, Antonis, & Reynolds, 1972; Duncan, Martens, & Ward, 1997; Keitel, Maess, Schröger, & Müller, 2013) or when target processing and distractor processing are subject to different domain-specific capacity limits. We prefer the latter explanation, because it is more in line with the finding that voice adaptors are filtered out in the presence of another voice (Zäske, Fritz & Schweinberger, 2013). It also parallels reports that irrelevant face processing is reduced under high load when attending another target face, but not when attending other target objects, such as houses or hands (Neumann, Mohamed, & Schweinberger, 2009, 2011), or letter strings, as in the standard perceptual-load task (Neumann & Schweinberger, 2008). Importantly, the present results are therefore not necessarily inconsistent with studies showing effects of visual perceptual load on auditory processing (Macdonald & Lavie, 2011; Molloy et al., 2015; Raveh & Lavie, 2015), as these studies used simple tones rather than voices as the task-irrelevant stimuli.

Our findings are potentially related to evidence that the duration of visual motion aftereffects is also unaltered by auditory perceptual load (Rees et al., 2001), and to electrophysiological data that the mismatch negativity (MMN) to rare frequency or intensity changes of task-irrelevant tone pips is unaffected by the difficulty of a concurrent visual discrimination task (Muller-Gass, Stelmack, & Campbell, 2006). Future research will be needed to establish in more detail how attentional resources are allocated both between modalities and between specific stimulus domains within modalities. Specifically, we expect that systematic manipulation of different auditory target and distractor domains in the context of selective attention tasks will contribute more detailed information with respect to the existence of domain-specific attentional resources for certain kinds of auditory stimuli (e.g., human voices, similar to what has been proposed in the visual modality for faces).

The finding of a significant VGAE in the absence of attention to voice adaptors suggests that despite being irrelevant for the task at hand, voice gender was sufficiently processed for adaptation to occur. This is in line with the notion that audition is an "early-warning" system (Murphy et al., 2013) that constantly processes auditory input, independent of the attentional focus. However, this may seem at odds with our previous finding that the VGAE is abolished when gender-specific voice adaptors are ignored during dichotic adaptation in the presence of simultaneous androgynous adaptor voices in the attended ear (Zäske, Fritz, & Schweinberger, 2013). A possible explanation for this discrepancy may be the existence of voice-specific attentional resources, as discussed above. Accordingly, in Zäske, Fritz, and Schweinberger's study, attention to an androgynous voice may have exhausted voice-specific resources, leaving little or no capacity for the processing of irrelevant gender-specific voice adaptors. The voice adaptors in the present study, by contrast, were presented along with task-relevant alphanumeric characters, which presumably spared voice-specific attentional resources, and thereby preserved processing of the voice adaptors (for a similar argument for faces, see Bindemann, Burton, & Jenkins, 2005; Neumann et al., 2011; Neumann & Schweinberger, 2008, 2009).

Note that the test stimuli were overall perceived as slightly more male than female. Although one may expect that morphed voices that are physically intermediate between original male and female speakers should on average be perceived equally often as male and female, deviations from 50 % male/female classifications are common in research using gender-morph continua, both with and without adaptation (e.g., Skuk & Schweinberger, 2014; Zäske, Fritz, & Schweinberger, 2013; Zäske, Schweinberger, Kaufmann, & Kawahara, 2009; Zäske, Skuk, et al., 2013). The present asymmetry could therefore reflect stronger aftereffects following female than following male voice adaptors, or may be the result of a general bias to perceive voices as "male." The latter notion could be related to recent findings that the perception of gender-morphed voices can vary substantially between listeners and between speaker identities (Skuk, Dammann, & Schweinberger, 2015). To assess whether the VGAE is susceptible to visual working memory load, we conducted a second experiment. In Experiment 2, the participants adapted to irrelevant voices while recognizing digits they had previously encountered. This was done to test whether high working memory load would *increase* the processing of adaptor voices (Lavie et al., 2004). Alternatively, the processing of adaptor voices could be unaffected by the load of a crossmodal working memory task, similar to findings in the face domain (Moradi et al., 2005).

Experiment 2

Method

Participants Thirty-two student participants (mean age = 22.5 yrs; range: 18–31; 16 female, one left-handed) contributed data. All of the participants were native German speakers, and none was familiar with any of the voices or reported hearing problems. All participants gave written informed consent and received course credit and an additional performance-based incentive of €1 or €2. The study was conducted in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of Friedrich Schiller University. None of the participants had taken part in Experiment 1.

Stimuli The voice stimuli were identical to those in Experiment 1. The single digits presented in the visual working memory task subtended a visual angle of 0.44° × 0.71°.

Procedure The procedures were analogous to those of Experiment 1, with the exception of the visual memory task described below (also see Fig. 1). On each trial, participants performed a visual working memory task while hearing three irrelevant adaptor voices. Specifically, participants had to remember six consecutive digits (0–9) that preceded the three male or female voice adaptors. To manipulate working memory load, the digits were either identical (low load) or all different (high load). During voice adaptation, the participants then classified several consecutive test digits as "studied" or "novel." Trials started with a red fixation cross for 1,000 ms, followed by alternating presentation of a study digit (500 ms) and a black screen (100 ms). After the sixth study digit, a 1,000-ms backward mask (#) announced the upcoming test. The onset of the first test digit coincided with the onset of the first of the three identical voice adaptors. Using the "d" and "l" keys, participants indicated as quickly and as accurately as possible whether or not a given test digit had just been presented. The digits were randomly generated such that studied digits appeared with a 60 % probability at test. As soon as a response was entered, a new test digit appeared, and so on until the offset of the third voice adaptor. Following the offset of the third voice adaptor, the trial procedure was identical to that of Experiment 1 (see Fig. 1), such that the adaptation and test phases had the same timing and duration (9,444 ms) in both experiments. Including the encoding phase of the working memory task (4,500 ms), the overall trial duration was 13,944 ms. Overall, Experiment 2 lasted ~30 min.

Table 2 Mean performance (*M*) and standard deviations (*SD*) in the selective attention task for low and high visual working memory load, depicted separately for the mean number of displays (correct displays/total number), accuracy, and reaction times (RTs) for correctly classified displays

Level of Load	Displays (*n*)		Accuracy (%)		Correct RT (ms)	
	M	*SD*	*M*	*SD*	*M*	*SD*
Low	222/235	27	94.6	3.6	437	55
High	146/176	25	82.7	7.0	619	110

Results

Validation of the load manipulation The successful manipulation of load was confirmed by ANOVAs with repeated measures on level of load (low/high) conducted for all performance measures in the selective attention task (Table 2): more correct displays under low than under high load [$F(1, 31) = 731.02$, $p <$ .001, η_p^2 = .959], faster correct RTs during low than during high load [$F(1, 31) = 220.35$, $p <$ .001, η_p^2 = .877], and more correct responses during low than during high load (*M* = 94.6 % vs. *M* = 82.7 %) [$F(1, 31) = 111.66$, $p <$ .001, η_p^2 = .783].

Voice gender aftereffects An ANOVA on the proportions of "female" responses to androgynous test voices with repeated measures on adaptation condition (female/male) and level of load (low/high) revealed a significant VGAE, with more "female" responses following male than following female adaptation [$F(1, 31) = 41.39$, $p <$ .001, η_p^2 = .572] (*M* = 61.6 % and 44.6 % "female" responses, respectively; see Fig. 3). An interaction of adaptation condition and level of load [$F(1, 31) = 5.04$, p = .032, η_p^2 = .140] reflected a larger

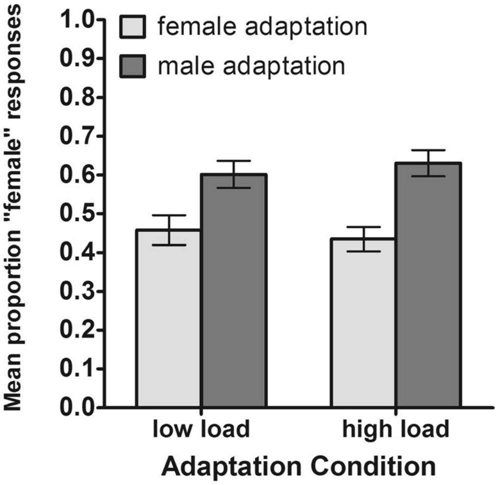

Fig. 3 Mean proportions of "female" responses following adaptation to female and male voices under low and high load in a visual working memory task (Exp. 2). Note that the voice aftereffects are increased under high relative to low visual working memory load. Error bars indicate *SEM*s

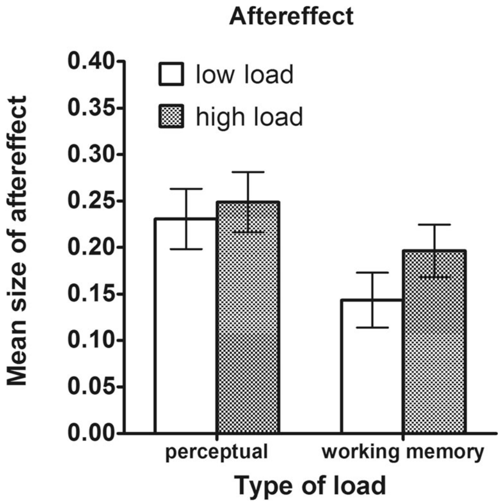

Fig. 4 Mean sizes of the voice gender aftereffects (male minus female adaptation) after low and high perceptual load (Exp. 1) and working memory load (Exp. 2). Error bars indicate *SEM*s

VGAE under high than under low working memory load (M_{diff} = 19.6 % vs. 14.3 % "female" responses, respectively). Please refer to Fig. 4 for the mean sizes of the aftereffects in Experiments 1 and 2.

Discussion

We found significant VGAEs under both low- and high-load conditions, suggesting that despite being irrelevant for the task at hand, voice gender was sufficiently processed for adaptation to occur. Importantly, and in line with load theory (Lavie et al., 2004), the VGAE was increased under high visual working memory load (Exp. 2). Accordingly, high load disrupted stimulus-processing priorities, allowing task-irrelevant voice adaptors to be processed to a larger extent than under low load. Since executive control over task priorities is a high-level cognitive function, these results support the notion that voice adaptation occurs at higher-level processing stages (Schweinberger et al., 2008; Zäske, Fritz, & Schweinberger, 2013). These findings pose an interesting contrast to linguistic aftereffects of speech adaptation, which do not appear to depend on visual working memory and which proceed automatically (Baart & Vroomen, 2010; Samuel & Kat, 1998). How do these findings relate to reports that the face identity aftereffect is not susceptible to an auditory as opposed to a visual working memory task (Moradi et al., 2005)? A tentative explanation may be that intermodal attentional resources are asymmetrically distributed between vision and audition, causing visual working memory tasks to have a higher impact on auditory distractor processing than in the opposite direction. In this context, the possible role of phonological recoding of the visual stimuli for the present working memory task may deserve particular consideration.

General discussion

Here we demonstrated that voice processing, as reflected in the voice gender aftereffect, is preserved despite selective attention to visual tasks during voice adaptation. Importantly, whereas the magnitude of the VGAE was increased under high relative to low working memory load (Exp. 2), in line with load theory (Lavie et al., 2004), the VGAE was completely unaffected by perceptual load (Exp. 1), at variance with load theory.

Taken together, the present results highlight limitations to the automaticity of voice processing (Zäske, Fritz, & Schweinberger, 2013), thereby pointing to an important difference from more "automatic" linguistic aftereffects (Baart & Vroomen, 2010; Samuel & Kat, 1998). Since the processing of unattended voices is enhanced by high visual working memory load, we suggest that voice adaptation occurs at higher-level processing stages, for which memory load effects would occur independently of target and distractor domains or modalities. By contrast, effects of perceptual load on voice processing depend on the domain of the target stimuli, and thus reflect domain-specific capacity limits. In conclusion, working memory, but not general perceptual capacities, determines the extent of voice processing during visual selective attention.

Author note We thank Sascha Müller for assistance with the experimental programming, and Dana Schneider for inspiring discussions. We further thank three anonymous reviewers for helpful comments, and the Deutsche Forschungsgemeinschaft (DFG, ZA 745/1-2) and the Friedrich Schiller University (Amélie Mummendey Young Researchers' Award) for financial support.

References

Alain, C., & Izenberg, A. (2003). Effects of attentional load on auditory scene analysis. *Journal of Cognitive Neuroscience, 15,* 1063–1073.

Allport, D. A., Antonis, B., & Reynolds, P. (1972). On the division of attention: A disproof of the single channel hypothesis. *Quarterly Journal of Experimental Psychology, 24,* 225–235.

Baart, M., & Vroomen, J. (2010). Phonetic recalibration does not depend on working memory. *Experimental Brain Research, 203,* 575–582.

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

Berman, R. A., & Colby, C. L. (2002). Auditory and visual attention modulate motion processing in area MT+. *Cognitive Brain Research, 14,* 64–74.

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

Bindemann, M., Burton, A. M., & Jenkins, R. (2005). Capacity limits for face processing. *Cognition, 98,* 177–197.

Brand-D'Abrescia, M., & Lavie, N. (2008). Task coordination between and within sensory modalities: Effects on distraction. *Perception & Psychophysics, 70,* 508–515. doi:10.3758/PP.70.3.508

Conway, A. R. A., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin & Review, 8,* 331–335. doi:10.3758/BF03196169

Dalton, P., Santangelo, V., & Spence, C. (2009). The role of working memory in auditory selective attention. *Quarterly Journal of Experimental Psychology, 62,* 2126–2132.

de Fockert, J. W. (2013). Beyond perceptual load and dilution: A review of the role of working memory in selective attention. *Frontiers in Psychology, 4,* 287. doi:10.3389/fpsyg.2013.00287

Duncan, J., Martens, S., & Ward, R. (1997). Restricted attentional capacity within but not between sensory modalities. *Nature, 387,* 808–810. doi:10.1038/42947

Eimas, P. D., & Corbit, J. D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology, 4,* 99–109.

Fairnie, J., Moore, B. C. J., & Remington, A. (2016). Missing a trick: Auditory load modulates conscious awareness in audition. *Journal of Experimental Psychology: Human Perception and Performance*. doi:10.1037/xhp0000204. **Advance online publication**.

Jacoby, O., Hall, S. E., & Mattingley, J. B. (2012). A crossmodal crossover: Opposite effects of visual and auditory perceptual load on steady-state evoked potentials to irrelevant visual stimuli. *NeuroImage, 61,* 1050–1058.

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., & Matsui, H. (2003). Auditory morphing based on an elastic perceptual distance metric in an interference-free time-frequency representation. In *Proceedings of the 2003 I.E. International Conference on Acoustics, Speech, and Signal Processing: Vol. I* (pp. 256–259). Piscataway, NJ: IEEE Press.

Keitel, C., Maess, B., Schröger, E., & Müller, M. M. (2013). Early visual and auditory processing rely on modality-specific attentional resources. *NeuroImage, 70,* 240–249.

Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in Cognitive Sciences, 9,* 75–82. doi:10.1016/j.tics.2004.12.004

Lavie, N., Hirst, A., de Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General, 133,* 339–354. doi:10.1037/0096-3445.133.3.339

Macdonald, J. S. P., & Lavie, N. (2011). Visual perceptual load induces inattentional deafness. *Attention, Perception, & Psychophysics, 73,* 1780–1789. doi:10.3758/s13414-011-0144-4

Molloy, K., Griffiths, T. D., Chait, M., & Lavie, N. (2015). Inattentional deafness: Visual load leads to time-specific suppression of auditory evoked responses. *Journal of Neuroscience, 35,* 16046–16054.

Moradi, F., Koch, C., & Shimojo, S. (2005). Face adaptation depends on seeing the face. *Neuron, 45,* 169–175. doi:10.1016/j.neuron.2004.12.018

Mullennix, J. W. (1986). *Attentional limitations in the perception of speech.* Unpublished doctoral dissertation, State University of New York, Buffalo, NY.

Muller-Gass, A., & Schröger, E. (2007). Perceptual and cognitive task difficulty has differential effects on auditory distraction. *Brain Research, 1136,* 169–177.

Muller-Gass, A., Stelmack, R. M., & Campbell, K. B. (2006). The effect of visual task difficulty and attentional direction on the detection of acoustic change as indexed by the Mismatch Negativity. *Brain Research, 1078,* 112–130.

Murphy, S., Fraenkel, N., & Dalton, P. (2013). Perceptual load does not modulate auditory distractor processing. *Cognition, 129,* 345–355.

Neumann, M. F., Mohamed, T. N., & Schweinberger, S. R. (2009). Preserved encoding of unfamiliar faces under high attentional load: ERP evidence. *Psychophysiology, 46,* S134–S135.

Neumann, M. F., Mohamed, T. N., & Schweinberger, S. R. (2011). Face and object encoding under perceptual load: ERP evidence. *NeuroImage, 54,* 3021–3027. doi:10.1016/j.neuroimage.2010.10.075

Neumann, M. F., & Schweinberger, S. R. (2008). N250r and N400 ERP correlates of immediate famous face repetition are independent of perceptual load. *Brain Research, 1239,* 181–190. doi:10.1016/j.brainres.2008.08.039

Neumann, M. F., & Schweinberger, S. R. (2009). N250r ERP repetition effects from distractor faces when attending to another face under load: Evidence for a face attention resource. *Brain Research, 1270,* 64–77.

Raveh, D., & Lavie, N. (2015). Load-induced inattentional deafness. *Attention, Perception, & Psychophysics, 77,* 483–492.

Rees, G., Frith, C., & Lavie, N. (2001). Processing of irrelevant visual motion during performance of an auditory attention task. *Neuropsychologia, 39,* 937–949. doi:10.1016/S0028-3932(01)00016-1

Samuel, A. G., & Kat, D. (1998). Adaptation is automatic. *Perception & Psychophysics, 60,* 503–510.

Schweinberger, S. R., Casper, C., Hauthal, N., Kaufmann, J. M., Kawahara, H., Kloth, N., . . . Zäske, R. (2008). Auditory adaptation in voice perception. *Current Biology*, *18*, 684–688. doi:10.1016/j.cub.2008.04.015

Schweinberger, S. R., Kawahara, H., Simpson, A. P., Skuk, V. G., & Zäske, R. (2014). Speaker perception. *Wiley Interdisciplinary Reviews: Cognitive Science, 5,* 15–25.

Skuk, V. G., Dammann, L. M., & Schweinberger, S. R. (2015). Role of timbre and fundamental frequency in voice gender adaptation. *Journal of the Acoustical Society of America, 138,* 1180–1193.

Skuk, V. G., & Schweinberger, S. R. (2013a). Adaptation aftereffects in vocal emotion perception elicited by expressive faces and voices. *PLoS ONE, 8*, e81691. doi:10.1371/journal.pone.0081691

Skuk, V. G., & Schweinberger, S. R. (2013b). Gender differences in familiar voice identification. *Hearing Research, 295,* 131–140.

Skuk, V. G., & Schweinberger, S. R. (2014). Influences of fundamental frequency, formant frequencies, aperiodicity, and spectrum level on the perception of voice gender. *Journal of Speech, Language, and Hearing Research, 57,* 285–296. doi:10.1044/1092-4388(2013/12-0314)

Sussman, J. E. (1993). Focused attention during selective adaptation along a place of articulation continuum. *Journal of the Acoustical Society of America, 93,* 488–498.

Tellinghuisen, D. J., & Nowak, E. J. (2003). The inability to ignore auditory distractors as a function of visual task perceptual load. *Perception & Psychophysics, 65,* 817–828.

Webster, M. A., & MacLeod, D. I. A. (2011). Visual adaptation and face perception. *Philosophical Transactions of the Royal Society B, 366,* 1702–1725.

Zäske, R., Fritz, C., & Schweinberger, S. R. (2013). Spatial inattention abolishes voice adaptation. *Attention, Perception, & Psychophysics, 75,* 603–613. doi:10.3758/s13414-012-0420-y

Zäske, R., Schweinberger, S. R., Kaufmann, J. M., & Kawahara, H. (2009). In the ear of the beholder: Neural correlates of adaptation to voice gender. *European Journal of Neuroscience, 30,* 527–534. doi:10.1111/j.1460-9568.2009.06839.x

Zäske, R., Schweinberger, S. R., & Kawahara, H. (2010). Voice aftereffects of adaptation to speaker identity. *Hearing Research, 268,* 38–45. doi:10.1016/j.heares.2010.04.011

Zäske, R., Skuk, V. G., Kaufmann, J. M., & Schweinberger, S. R. (2013). Perceiving vocal age and gender: An adaptation approach. *Acta Psychologica, 144,* 583–593. doi:10.1016/j.actpsy.2013.09.009

Zäske, R., Volberg, G., Kovács, G., & Schweinberger, S. R. (2014). Electrophysiological correlates of voice learning and recognition. *Journal of Neuroscience, 34,* 10821–10831. doi:10.1523/JNEUROSCI.0581-14.2014