On the Representation of Speaker Information in Human Voices: An Adaptation Approach.

Dissertation
zur Erlangung des akademischen Grades
doctor philosophiae (Dr. phil.)

vorgelegt dem Rat der Fakultät für Sozial- und Verhaltenswissenschaften
der Friedrich-Schiller-Universität Jena
von Dipl.-Psych. Romi Zäske
geboren am 22.06.1981 in Köthen (Anhalt)

Gutachter

1. Prof. Dr. Stefan R. Schweinberger (Friedrich-Schiller-Universität Jena)
2. Prof. Dr. Pascal Belin (University of Glasgow)

Tag der mündlichen Prüfung: 14.10.2010

Table of Contents

Preface

Human voices are not only transmitters of speech but also carry important *paralinguistic* speaker information that helps guide daily interactions. Anatomical differences in the vocal apparatus as reflected in unique acoustic patterns make voices of individuals and groups of speakers distinguishable and recognisable. Whereas voice perception abilities may be similarly important for efficient social interactions as are speech or face perception abilities, relatively little effort has been undertaken to systematically study these processes. This has been attributed to the fragmentation of voice research across many disciplines and the primary focus on practical issues. These circumstances not only complicated the monitoring of empirical evidence and the exchange of knowledge but also hampered the establishment of theoretical models (Kreiman, 1997).

At present, voice perception is rejoicing increasing scientific attention. This has been mainly prompted by theoretical input from face research and advancements in brain imaging technology. Only ten years ago functional magnetic resonance imaging (fMRI) started to elucidate brain regions selectively involved in voice processing (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000) thereby complementing earlier reports of a specialised neural substrate of face perception (Kanwisher, McDermott, & Chun, 1997). Moreover, considerable behavioural and neurophysiological evidence suggests that face and voice recognition systems interact at some stage (for reviews see Belin, Fecteau, & Bedard, 2004; Campanella & Belin, 2007). Therefore, the most influential models of person perception today, take an integrative neurocognitive perspective by assuming broadly analogous pathways for the processing of faces and voices via hierarchically organised modules (Belin et al., 2004; Neuner & Schweinberger, 2000).

Conceptual overlap often entails similar research questions and methodological approaches. Accordingly, the present work was inspired by the late discovery of high-level aftereffects in face perception following adaptation. These have been informative regarding the mental representation of various social signals in faces (Leopold, O'Toole, Vetter, & Blanz, 2001; Webster, Kaping, Mizokami, & Duhamel, 2004). In analogy, the present studies were designed to investigate auditory adaptation effects in voice perception which would allow for conclusions as to how certain aspects of voice quality are coded in memory.

The investigation of face aftereffects critically depends on sophisticated image morphing, such as stimulus manipulation techniques, by means of which facial features can be systematically distorted along arbitrary feature dimensions. A recent auditory analogue to this software (Kawahara & Matsui, 2003) finally provides a tool to create voice morphs, i.e.

interpolations between any two voices. Hence, voice morphs systematically vary with respect to all social signals contained in the original samples, making them extremely valuable for psychophysical and neurocognitive research.

The following investigations are based on this novel technique. Using behavioural and electrophysiological measures in the context of adaptation paradigms the present research intends to broaden our sparse knowledge on the neural representations of voice gender and speaker identity. The three studies reported here have either been already published (studies I and II) or accepted for publication (study III) in scientific journals. While only these three studies are covered in detail as part of my dissertation, I have also included preliminary reports of additional and as yet unpublished data from two further experiments (chapter 4). Following a discussion of the present results within a working model of voice perception, I will offer some suggestions for future research.

I would like to cordially thank Stefan Schweinberger whose professional support as my supervisor and whose unequalled enthusiasm enriched my work and made my research both fruitful and enjoyable. It has been and is continuing to be a great pleasure to work with him and as part of this splendid team. Further, I am much obliged to Jürgen Kaufmann who first introduced me to the field of person perception and who has become a dear colleague and advisor. Finally yet importantly, my Mum deserves a great hug for her encouragement and her constant belief in me.

Romi Zäske
Jena, 22nd June 2010

1 Social Information in Voices

Acoustic Correlates

Voices are complex acoustic signals which, like any other sound, are characterised by changes in air pressure. Vocal signals are composed of their fundamental frequency (F0) as well as formant frequencies (F1, F2, …Fn) which are integers of F0 with relatively high sound pressure levels. According to source-filter theory (Fant, 1960, as cited in Belin et al., 2004; Fitch, 2000), F0 is determined by the vibration of the vocal cords (i.e. the source) as caused by air flowing from the lungs through the larynx. The resultant signal is then filtered by the supra-laryngeal vocal tract whose resonant characteristics determine the formant frequencies. F0 is perceived as voice pitch and can be identical for different speakers. The formant frequencies, by contrast, determine the unique timber of each voice and are largely independent of F0.

Intra- and interindividual differences in the anatomical makeup of the vocal apparatus and the way it is used, affect acoustic parameters that simultaneously code speech information and various social signals such as identity, gender, age, or emotional state. Hence, vocal patterns are subject to an enormous and yet systematic variability both within speakers as well as between individuals and groups of speakers. For instance, it is well known, that not only individual voices but also male and female voices differ with respect to F0 which is typically lower in men than in women (Klatt & Klatt, 1990). In addition, women's voices are breathier than men's are. At the same time, vocal parameters also underlie age-related changes some of which are similar for both genders: F0 standard deviation and breathiness increase whereas speech rate decreases in the elderly (Linville, 1996). Other parameters change as a function of gender. While F0 rises for men between middle and old age, it tends to fall for women (Torre & Barlow, 2009). To complicate matters, most of these differences are due to physiological changes, whereas others may be of functional origin. For instance, despite being physiologically capable of closing their vocal cords, young women may fail to do so in order to achieve a breathy voice quality (Linville, 1992; cited in Linville, 1996). Moreover, gender and age related vocal characteristics may be more or less pronounced in different individuals.

It is conceivable from these examples that numerous physiological and functional sources of intra- and interindividual variability, interactively contribute to the typical vocal patterns of particular speakers or groups of speakers (e.g. young men). Furthermore, the extraction of social signals such as gender and age at least partially relies on the same acoustic cues. Note, however, that information physically available in the voice signal need not

necessarily be perceptually relevant (e.g. Bruckert, Lienard, Lacroix, Kreutzer, & Leboucher, 2006; Linville & Fisher, 1985).

The Role of Speech in Voice Perception

It has been suggested that in order to understand speech, talker variability in speech production needs to be disregarded, a process termed talker normalisation. Conversely, speech normalisation implies the extraction of speech-invariant information across different utterances of the same speaker or various speakers, hence enabling the listener to recognise speakers (e.g. Schweinberger, Herholz, & Stief, 1997). Traditionally, speech and speaker perception have been regarded as entirely independent mechanisms, an assumption that is supported by findings of functional double dissociations in neurological patients. For instance, Assal, Aubert and Buttet (1981, as cited in Belin et al., 2004) reported cases of aphasia with normal voice recognition, as well as patients with impaired voice recognition but preserved speech perception abilities. Adding to these findings, a recent case study (Garrido et al., 2009) presented a woman suffering from developmental "phonagnosia", i.e. the inability to recognise familiar voices. Nevertheless, she performed normally in speech perception tasks.

Since human communication is largely speech-based, voice recognition is predominantly achieved from speech signals. However, only minimal phonetic information as contained in isolated vowels is sufficient to classify voices as familiar (Beauchemin et al., 2006) or to determine voice gender (Lass, Hughes, Bowyer, Waters, & Bourne, 1976). To some extent, voice gender can also be recognised from non-speech vocalisations including laughs, cries or coughs (Fecteau, Armony, Joanette, & Belin, 2004). The matching of non-speech sounds (e.g. whistles and kisses) to mute videos of faces is, however, harder than for syllables (Schmid & Ziegler, 2006). This has been attributed to stronger crossmodal connections in the speech domain as opposed to the non-speech domain (Loh, Schmid, Deco, & Ziegler, 2010) and indicates that speech may play an important role in the integration of audiovisual speaker information.

Although it has been demonstrated that even severely modulated unintelligible speech, as created by time-reversal, may still carry perceptually relevant identity information (Bricker & Pruzansky, 1966), other studies suggest that speaker-specific cues in time-reversed speech rather act to facilitate implicit but not explicit voice recognition (Schweinberger, 2001; Vanlancker, Kreiman, & Emmorey, 1985a). Furthermore, research on the language familiarity effect showed that people can learn to recognise individual speakers from a language they do not understand, although voice recognition rates are consistently higher when speakers use the

listeners' mother-tongue (Goggin, Thompson, Strube, & Simental, 1991; Perrachione & Wong, 2007; Strube, Goggin, & Thompson, 1988; Thompson, 1987).

In sum, these examples suggest that although the extraction of social information from voices does not depend on speech or the listener's ability to understand it, speech may provide additional cues to speaker characteristics. It has even been proposed that phonetic cues alone, as contained in sinewave replicas of speech, can signal speaker identity despite lacking F0 and the natural timbre of voices (Remez, Fellowes, & Rubin, 1997; Sheffert, Pisoni, Fellowes, & Remez, 2002).

Earlier behavioural evidence for partially dependent processing of linguistic information and speaker attributes (Mullennix & Pisoni, 1990; Nygaard, Sommers, & Pisoni, 1994) has recently received support from neurophysiological research (Kaganovich, Francis, & Melara, 2006; Knösche, Lattner, Maess, Schauer, & Friederici, 2002). These studies suggest that the integration of phonetic and speaker information in voices may take place at an early (preattentive) level. However, interactions of speech and speaker processing were also observed at later, i.e. at semantic processing stages. Tesink and colleagues (2009), for instance, found that conflicting social information as conveyed by the speaker's voice (i.e. age, gender and social background) and sentence meaning is associated with enhanced activation in bilateral inferior frontal gyri. The authors related this activation to processes of unifying conflicting aspects of an utterance and their integration into representations of current expectations.

Cues to talker characteristics have been referred to as *non-linguistic* (also paralinguistic, extralinguistic or indexical) in the sense that they are implicitly contained in the speech signal, but distinct from the content of the message (Pisoni, 1997). However, while the conceptual differentiation between talker attributes (Who is speaking) and speech content (What is said) is straightforward, a strict dissociation of their physical bases seems infeasible given that both sources of information are contained in the same signal. Researchers have therefore criticised the traditional practice of treating linguistic and non-linguistic properties of speech as separate entities (Pisoni, 1997). Following from these considerations and the findings reported above, the processing of speech and speaker information appears to be at least partially dependent.

1.1 Perceiving Voice Gender

The perception of voice gender is among the most intensively studied non-linguistic speaker attributes and has been indirectly addressed in speech perception research. For instance, the voice-connotation hypothesis (Geiselman & Bellezza, 1976; 1977) suggests that speaker gender is processed automatically as is it sometimes influences the meaning of what is said. The notion of automatic gender processing came from observations of unimpaired sentence recall when listeners had to remember the speaker's gender in addition to sentences. One study investigating the relationship between acoustics and perception in speaker recognition found that gender judgements from isolated vowels uttered by adults are primarily based on F0 and less so on formants (Lass et al., 1976) because accuracies in this study significantly dropped when voiced samples were replaced with whispered samples lacking F0 information. Similarly, Gelfer and Mikos (2005) showed that gender judgements of natural vowels synthesised to have typical male (120 Hz) or female (240 Hz) pitch are based on F0 information alone even if formant frequencies provide a contradicting cue to gender. Fecteau and others (2004) showed that implicit gender recognition from non-speech sounds depends on speaker gender as priming effects are more pronounced in female voices than in male voices. Additionally, talker gender may be more efficiently judged from male than from female vowels (Owren, Berkowitz, & Bachorowski, 2007).

Another group (Lattner, Meyer, & Friederici, 2005) manipulated pitch information in two-word sentences uttered by one male and one female talker thereby creating atypical voice qualities, i.e. a low-pitched female voice vs. a high-pitched male voice. From fMRI data the authors concluded that different aspects of speaker gender are predominantly processed in the right hemisphere. Specifically, they identified functionally segregated areas involved in the processing of (1) voice pitch (close and anterior to right primary auditory cortex), (2) voice spectral information (right posterior superior temporal gyrus (STG) and areas surrounding the planum parietale bilaterally) and (3) voice naturalness, i.e. prototypicality (right anterior STG). This interpretation might be limited by the fact that only one speaker from every gender group was used. One can therefore not exclude that activation differences reported in that study were due to different speaker identities rather than genders. However, an earlier study by the same group using the same stimulus manipulation technique demonstrated that listeners may have expectations as to the typical composition of male and female voices (Lattner et al., 2003). The violation of these expectations or "perceptual templates" in an oddball paradigm, so it was reasoned, caused to a preattentive mismatch response in the magnetoencephalogram (MEG) at 200 ms after voice onset. The magnitude of this response,

also termed mismatch negativity (MMN), was independent of the acoustic similarity of standards and deviant voices. Although the authors do not report how they controlled for the "prototypicality" of the original samples, these studies tentatively suggest the existence of long-term representations of voice gender.

Sokhi and others (2005) had male participants perform a gender classification task on pitch-scaled and unaltered sentences from 12 male and female speakers respectively. They found that the mesio-parietal precuneus area responded more to male voices than to female voices, whereas the reverse situation was observed for the right anterior temporal gyrus near STS. In line with these findings Lattner et al. (2005) also demonstrated stronger activation for female voices than for male voices in the right superior temporal cortex. As a possible explanation for this result, Lattner and others reasoned that female/high-pitched voices could be of greater biological or social relevance than male voices. Children, for instance, prefer female over male voices (Fifer and Moon, 1988, cited in Lattner et al., 2005) and adults' enhanced physical responses to female voices may be due to their resemblance with high-pitched children's voices.

1.2 Perceiving Voice Identity

The systems subserving voice perception abilities may be phylogenetically old ones. Many animals rely on them during mate selection, for the recognition of their offspring or for distinguishing among other members of their social group (Bee & Gerhardt, 2002; Mathevon, Charrier, & Aubin, 2004; Rendall, Rodman, & Emond, 1996; e.g. Torriani, Vannoni, & McElligott, 2006). In many species including humans, the ability to recognise individual voices may be innate as newborns can instantly recognise their parents' voices (Decasper & Fifer, 1980; Ockleford, Vince, Layton, & Reader, 1988). Moreover, in humans as well as in non-human primates, the neural substrates of such abilities have recently been uncovered by brain imaging research (for recent reviews see Belin, 2006; Belin et al., 2004; Petkov, Logothetis, & Obleser, 2009; Watson, 2009).

Voice recognition, identification and discrimination are terms which have been used in an inconsistent manner although there is considerable evidence to suggest that these abilities depend on partially dissociable cognitive processes and brain structures (Garrido et al., 2009; Vanlancker & Kreiman, 1987; von Kriegstein & Giraud, 2004; Warren, Scott, Price, & Griffiths, 2006). In accordance with traditional models of face perception (Bruce & Young, 1986; Burton, Bruce, & Johnston, 1990), speaker recognition in the present work refers to situations when hearing a voice or seeing a face elicits a feeling of familiarity. This is independent of whether or not we have or have access to stored semantic information on the

speaker. The retrieval of this sort of information (e.g. occupation, name, etc.) from long-term memory will be termed identification and can be achieved only for voices *known* to a listener. By contrast, similarity judgements or discrimination tasks can be performed for known and unknown voices.

It has been suggested that absolute formant frequencies indicate individual speakers whereas their relative differences indicate vowels (Hirahara & Kato, 1992, as cited in Pisoni, 1997). This is supported by studies demonstrating that the discrimination between unfamiliar voices and the recognition of familiar voices largely relies on spectral information. For instance, multidimensional analyses on similarity ratings for vowels uttered by unfamiliar speakers (Baumann & Belin, 2010) suggested that F0 and formant frequencies best accounted for the variance. Furthermore, repetition priming for familiar (i.e. famous) voices was observed even when primes were time-reversed speech samples with distorted phonetic and articulatory cues but preserved frequency characteristics (Schweinberger, 2001).

However, experiments with backward and rate-altered speech from celebrities showed that listeners use a different set of acoustic parameters (i.e. patterns) for each speaker they recognise (Vanlancker et al., 1985a; Vanlancker, Kreiman, & Wickens, 1985b). Similarly, while the identification of personally familiar speakers from resynthesised vowels seems to dependent more on vocal tract features (i.e. formants) than on the glottal waveform on average, the relative contribution of these features to recognition varies between speakers (Lavner, Gath, & Rosenhouse, 2000) and listeners for natural vowels (Lavner, Rosenhouse, & Gath, 2001).

The Role of Voice Familiarity

Largely motivated by practical considerations, early voice perception research was focussed on uncovering factors related to the recognition of unfamiliar voices. This was in order to assess or improve the reliability of earwitness testimony in the forensic context. Meanwhile, cognitive psychologists emphasise the perception of familiar voices in analogy to face recognition research and further distinguish between famous and personally familiar voices. This classification is justified for various reasons. While the recognition of unfamiliar voices is highly unreliable (for a review see Clifford, 1980) people are usually more accurate at recognizing familiar speakers even from very short utterances such as vowels (Schmidtnielsen & Stern, 1985). Note, however, that even for personally familiar voices, recognition rates may be subject to extreme variability both between speakers and listeners, at least when based on isolated vowels (Lavner et al., 2001). Moreover, familiarity with a voice

can facilitate stream segregation (Newman & Evers, 2007) and the recognition of speech (Nygaard & Pisoni, 1998; Sheffert & Olson, 2004). This may be related to the notion that voices, and personally familiar ones in particular, capture attention (Beauchemin et al., 2006; Levy, Granot, & Bentin, 2001; Levy, Granot, & Bentin, 2003).

Studies with brain-damaged patients showed that impairments in the recognition of familiar voices, i.e. phonagnosia, are related to right-hemispheric damage whereas difficulties with the discrimination of unfamiliar voices occur with lesions in either hemisphere (Vanlancker & Kreiman, 1987; Vanlancker, Cummings, Kreiman, & Dobkin, 1988). In healthy listeners, familiar and unfamiliar voices elicit different cortical activation patterns as evidenced by fMRI studies. Specifically, whereas voices relative to non-vocal sounds selectively engage the right anterior STS (Belin, Zatorre, & Ahad, 2002; Belin et al., 2000) irrespective of familiarity (von Kriegstein & Giraud, 2004), von Kriegstein and Giraud found that the right posterior STS responded more strongly to unfamiliar voices than to personally familiar voices. Conversely, auditory cortex in left middle temporal gyrus (MTG) is preferentially activated by personally familiar voices possibly facilitating speech processing (Birkett et al., 2007). Data from positron emission tomography (PET) suggest that the correct recognition of personally familiar voices from 2 s sentences is associated with activation in the left frontal pole and the right temporal pole. This led to the conclusion that these areas contribute to the matching of currently heard voices and voice identity representations in memory (Nakamura et al., 2001). Furthermore, personally familiar voices (and faces) relative to unfamiliar ones increase activation in structures implicated in episodic memory and emotional salience, i.e. posterior cingulate cortex including retrosplenial cortex (Shah et al., 2001).

Furthermore, electrophysiological data suggest that the discrimination of human voices from other complex sounds starts as early as 164 ms following sound onset and culminates in a fronto-temporal positivity which is maximal at around 200 ms (Charest et al., 2009). Interestingly, a similar time window was reported for the preattentive detection of personally familiar voices (Beauchemin et al., 2006). Using an oddball paradigm, Beauchemin and others showed that 200 ms after stimulus onset, rare familiar voice deviants elicited a larger MMN than rare unfamiliar voice deviants within an unattended stream of an unfamiliar standard voice. Similarly, discrimination among unfamiliar voices seems to proceed early and automatically with the magnitude of mismatch responses varying as a function of voice similarity between standards and deviants (Titova & Näätänen, 2001).

The distinction of familiar and unfamiliar voice processing is reflected in voice perception models, which have been designed to primarily or exclusively explain one or the other. For instance, Papcun and colleagues' (1989) model proposes that unfamiliar speaker recognition is based on hard-to-remember voices representing a voice prototype. A prototype model has also been put forward to account for the perception of personally familiar male voices (Lavner et al., 2001). Van Lancker's et al. pattern recognition model (Vanlancker et al., 1985a; Vanlancker et al., 1985b) and Belin's neurocognitive model of voice perception (Belin et al., 2004) to which I will turn later in detail, are mainly concerned with familiar voice recognition and identification. However, none of the models that I know of addresses the process of familiarisation, i.e. voice learning.

1.3 Perceiving Voice Age

The perception of vocal age information is still poorly understood, and relevant research was strongly biased towards using male voices (e.g. Hamsberger, Shrivastav, Brown, Rothman, & Hollien, 2008; Hartman & Danhauer, 1975; Ryan & Burk, 1974; Shipp & Hollien, 1969; Shipp & Hollien, 1972; Shipp, Qi, Huntley, & Hollien, 1992) with some exceptions (e.g. Linville & Fisher, 1985; Linville & Korabic, 1986). However, some behavioural studies suggest that age estimations for young vs. older adult voices rely more strongly on F0 than on formants (Linville, 1996). According to Linville, age estimations are also affected by speaking rate and breathiness in male speakers as well as differences in F0 standard deviation for both speaker genders. A recent study independently manipulated values of glottal pulse rate (F0) and vocal tract length (formants) in male vowels by means of auditory morphing. Combined age/gender judgements (boy/girl/man/woman) were found to equally depend on both kinds of information (Smith & Patterson, 2005). A follow-up study by the same authors aligned the range of F0 and formants in vowels spoken by one juvenile and adult male and female speaker respectively. While age of the original speakers could be discriminated despite these manipulations, listeners found it harder to judge the original gender (Smith, Walters, & Patterson, 2007). These findings indicate that the perception of speaker age (child vs. adult) from isolated, sustained vowels is not completely disrupted by spectral alterations and may additionally rely on other cues. Moreover, according to Linville (1996), the accuracy of age estimations depends on factors such as the number of response categories, the age and gender of the listener, and the length and kind of stimulus material (e.g. whispered vs. phonated vowels or sentences).

2 A Neurocognitive Framework of Voice Perception

Based on well-established models of person perception (Bruce & Young, 1986; Burton et al., 1990) analogous processing routes have been proposed for faces and voices (Belin et al., 2004; Campanella & Belin, 2007; Ellis, Jones, & Mosdell, 1997; Neuner & Schweinberger, 2000; e.g. Schweinberger et al., 1997). According to these models (see Figure 1 for a recent example), both stimuli initially undergo an analysis of their low-level features. Subsequently, extraction of features which are invariant across different conditions of presentation or production takes place (Neuner & Schweinberger, 2000). Following this stage of structural encoding three functionally independent pathways are assumed to mediate the recognition of affect, speech and identity information in each modality (Belin et al., 2004). The perception of person identity proceeds with a comparison of incoming structural representations to long-term representations as stored in "unimodal" voice recognition units (VRUs) and face recognition units (FRUs). If a sufficient match between these representations is detected activation further spreads to post-perceptual person identity nodes (PINs). PINs are conceived as points of multimodal convergence which provide access to semantic information and thereby enable the identification of a person (Neuner & Schweinberger, 2000). Recognition of a face or voice as accompanied by a feeling of familiarity is thought to arise either at the stage of recognition units (Bruce & Young, 1986) or at the PIN level (Burton et al., 1990). As can be seen in Figure 1, perceptual links as well as post-perceptual relays in principle allow for crosstalk between face and voice modules. It is a matter of debate, however, if and how information is exchanged via these links (Campanella & Belin, 2007).

The assumptions of hierarchical modularity are well underpinned by neuroimaging results (e.g. Warren et al., 2006). For instance, structural encoding may be instantiated by temporal voice areas (TVAs) and the fusiform face area (FFA) respectively (Belin et al., 2000; Haxby, Hoffman, & Gobbini, 2000). Campanella and Belin (2007) argue that the recognition of familiar voices is implemented in areas close to TVAs, i.e. in the right anterior STS, which is a potential anatomical substrate of the VRUs (Belin et al., 2004). Alternatively, voice recognition, i.e. the decision of whether a voice is familiar, could involve more posterior regions of the STS (von Kriegstein & Giraud, 2004) or a network comprising the left frontal pole, the right temporal pole, the right entorhinal cortex, and the left precuneus (Nakamura et al., 2001). In that study, all of these regions differentially responded to familiar and unfamiliar voices. Finally, the retrosplenial cortex has been suggested as a possible site of PINs as it is activated both by familiar voices and faces alone, when compared to unfamiliar voices or faces (Campanella & Belin, 2007; Shah et al., 2001).

However, it has also been proposed that the integration of voice and face information takes place at an earlier, i.e. perceptual level, via direct links between TVAs and the FFA (von Kriegstein, Kleinschmidt, Sterzer, & Giraud, 2005).

None of these models explicitly incorporated the perception of age and gender information from voices, possibly because their emphasis lies on the recognition of familiar people for whom identity is largely confounded with age and gender. The recognition of gender and age from unfamiliar *faces*, however, may be achieved via visually derived semantic codes as initially proposed by Bruce and Young (1986). This reasoning may also apply to voices, such that unfamiliar voices contain acoustically derived semantic codes.

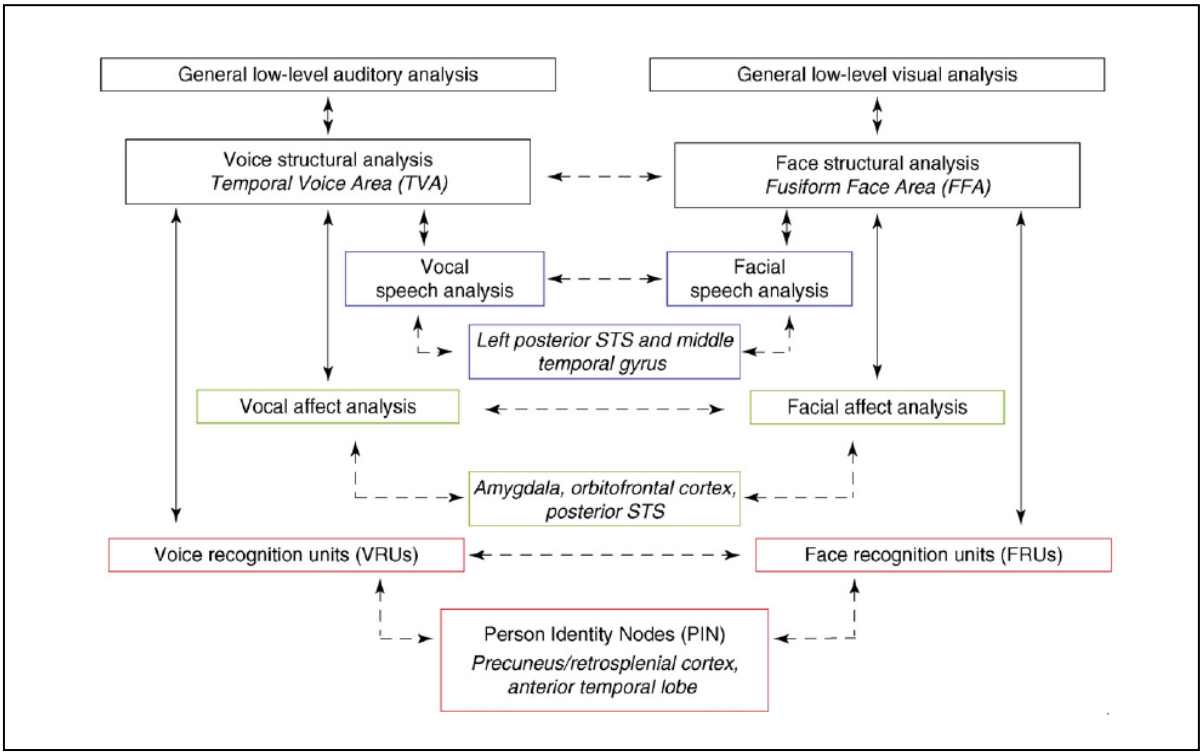


Fig. 1. Adaptation of Belin's et al. (2004) neurocognitive model of voice perception (from Campanella & Belin, 2007) as proposed in broad analogy to Bruce and Young's (1986) model of face perception. Processing pathways of voices and faces are depicted on the left and right side respectively. Dashed arrows indicate possible crosstalk between modalities via direct links (left-right arrows) and via supramodal regions of integration (L-shaped arrows).

3 The Present Studies

The previous chapters have outlined that the recognition of identity and gender information from voices probably relies on specialised structures predominantly located in the right hemisphere. Moreover, there is evidence that some voice characteristics, i.e. familiarity (Beauchemin et al., 2006) and identity of unfamiliar voices (Titova & Näätänen, 2001), are detected at an early preattentive stage. The exact coding mechanisms underlying the perception of voice characteristics such as gender and identity remain, however, unclear.

An Adaptation Approach

According to Clifford and Rhodes (2005) process-specific coding is a basic type of neural population coding that occurs when anatomically localized cell assemblies encode feature dimensions specifically required for the processing of particular objects such as faces (e.g. Kanwisher et al., 1997). They further suggest that the operation of specialized cell assemblies may enhance the discriminability of behaviourally-relevant stimuli and may have developed to " […] anatomically segregate computationally conflicting processes" (Clifford & Rhodes, 2005, p. 138). Following this reasoning, the need for process-specific coding may also exist in the auditory modality, as not only faces, but also voices are complex (auditory) objects carrying behaviourally relevant social signals. Not surprisingly therefore, fMRI research uncovered systems selectively tuned to respond to non-linguistic voice properties (Belin & Zatorre, 2003; Belin et al., 2000).

Recently, adaptation paradigms have been introduced to investigate high-level coding of facial signals. Adaptation, a mechanism by which sensory processing is continuously recalibrated according to incoming stimuli, is revealed by contrastive aftereffects. For instance, prolonged stimulation with a moving object such as a waterfall or a river causes static objects subsequently presented in the same visual field to be perceived as moving in the opposite direction. This motion aftereffect is an example of low-level adaptation to a simple stimulus property, i.e. visual motion, and has been attributed to an activity imbalance in antagonistic neuron populations coding upward and downward motion respectively (Anstis, Verstraten, & Mather, 1998). While transient unimodal and crossmodal aftereffects of adaptation have primarily been demonstrated for simple features including visual and auditory motion (Dong, Swindale, & Cynader, 1999; Kitagawa & Ichihara, 2002), pitch (Kay & Matthews, 1972), or loudness (Dange, Warm, Weiler, & Dember, 1993) to name but a few, perceptual biases following high-level adaptation to complex face stimuli have only been uncovered recently. Some examples for social signals inducing contrastive aftereffects are

gender (Kovacs et al., 2006; Webster et al., 2004), identity (Leopold et al., 2001), expression and ethnicity (Webster et al., 2004) or eye gaze (Jenkins, Beaver, & Calder, 2006; Schweinberger, Kloth, & Jenkins, 2007).

Reports of auditory adaptation to complex objects have almost exclusively focussed on linguistic properties of speech rather than on non-linguistic speaker characteristics (e.g. Eimas & Corbit, 1973; Jusczyk & Luce, 2002; Morse, Kass, & Turkienicz, 1976; but see Mullennix, Johnson, TopcuDurgun, & Farnsworth, 1995). In order to probe the coding principles of *non-linguistic* voice information, i.e. gender and identity, the present studies employ auditory adaptation paradigms. The general rationale of these paradigms is that differences in perception of test voices are measured as a function of previous adaptation to categories of a particular feature dimension (e.g. male or female voice gender). Modern auditory morphing software (Kawahara & Matsui, 2003) as based on STRAIGHT algorithm (Speech Transformation and Representation using Adaptive Interpolation of weiGHTed spectrum, Kawahara, Masuda-Katsuse, & de Cheveigne, 1999) plays a key role in this endeavour. This algorithm allows for the creation of naturalistic test voices varying on the test dimension by means of bilinear time frequency interpolation between two original voice samples (e.g. a male and a female voice). As a result, newly created voice morphs (e.g. androgynous voices) systematically differ with respect to the proportion of social signals conveyed by the original voices.

Repetition effects for simple and complex sounds have not only been observed on a psychophysical level but are also reflected in changes of neural responses (for review see Jääskeläinen, Ahveninen, Belliveau, Raij, & Sams, 2007). These effects have been measured for instance with fMRI (e.g. Belin & Zatorre, 2003; Robson, Dorosz, & Gore, 1998), electroencephalography (EEG) and MEG (e.g. Altmann et al., 2008; Budd, Barry, Gordon, Rennie, & Michie, 1998; May et al., 1999; Schweinberger, 2001). However, the exact mechanisms underlying neuronal adaptation in the visual domain are still debated (Grill-Spector, Henson, & Martin, 2006). An advantage of the EEG is its high temporal resolution by which can precisely capture the timing of cognitive processes. Therefore, in order to assess the timing of potential voice adaptation effects using a more direct measure of neuronal activity, the present thesis also contains a study on electrophysiological correlates of voice adaptation.

Aims and Hypotheses

Previous behavioural and neurophysiological studies had indicated that the brain is equipped with voice detectors possibly specialised in processing non-linguistic speaker information of male and female voice quality (Lattner et al., 2005; Sokhi, Hunter, Wilkinson, & Woodruff, 2005) and speaker identity (Belin & Zatorre, 2003). Based on observations that cortical neurons tuned to a particular stimulus property selectively decrease in responsiveness when repeatedly or continuously stimulated, my hypotheses were that a decrease in responsiveness would also occur in neurons coding different non-linguistic voice features including gender and identity. Prolonged exposure to male and female voices was predicted to cause contrastive aftereffects in the perception of voice gender as reflected behaviourally in study I (Schweinberger et al., 2008) as well as electrophysiologically in study II (Zäske, Schweinberger, Kaufmann, & Kawahara, 2009). Furthermore, I expected contrastive aftereffects of adaptation to voice identity in study III when adaptors are personally familiar voices (Zäske, Schweinberger, & Kawahara, 2010). In addition to unimodal voice adaptation, conditions of crossmodal adaptation to faces (studies I and III) and names (study I) tested the modality specificity of voice gender and identity aftereffects. I further probed in all studies whether aftereffects would survive at least a few minutes.

3.1 Auditory Adaptation in Voice Perception (Schweinberger et al., 2008)

Whereas contrastive aftereffects of adaptation have provided insight into the coding of visual information in faces (Jenkins et al., 2006; e.g. Webster et al., 2004) and linguistic information in speech (Eimas & Corbit, 1973; Holt, 2006), high-level aftereffects in the processing of non-linguistic auditory information await to be demonstrated. The first study of the present thesis therefore investigated if adaptation to unfamiliar voices (Exp. 1) would induce aftereffects in the perception of voice gender. In each trial participants listened to four consecutive voice adaptors of male, female or androgynous speakers uttering vowel-consonant vowel syllables (VCVs) and performed a two-alternative forced choice gender classification task on test voices immediately following adaptors. Test voices varied with respect to their female/male proportion between 20/80% and 80/20%, in steps of 10%. These voice morphs had been created by interpolating between male and female original voices using STRAIGHT auditory morphing software. While androgynous adaptors were morphs with equal gender proportions (50/50%), male and female adaptation conditions used original male (0/100%) and female voices (100/0%).

In Exp. 1, participants not only perceived voice gender according to the proportion of gender information in test voices but also exhibited the predicted voice gender aftereffect (VGAE), such that prolonged stimulation with female voices biased the perception of subsequent test voices as appearing more male and vice versa. This aftereffect was largest for intermediate morph levels of test voice, indicating that stimuli inducing the highest degree of response uncertainty were most susceptible to adaptation. Further, a shift in the category boundary of listeners' internal voice gender continuum towards the adapting gender was evident in points of subjective equality (PSE), i.e. test gender proportions for which responses are 50% male and 50% female. This shift was in the predicted direction for each but one participant, showing that the VGAE is a universal phenomenon. Further, it was demonstrated that despite being reduced after an adaptation-test-interval of at least a few minutes, the VGAE was still measurable at the most ambiguous morph level.

In order to test if the VGAE was due to adaptation to a more general gender concept rather than gender-specific auditory information, Exp. 2, which was otherwise identical with Exp.1, used written male or female first names as adaptors (with no androgynous adaptation condition). Unlike adaptation to voices in Exp. 1, adaptation to names did not elicit any voice aftereffects in a new group of participants. This result is broadly consistent with the lack of crossdomain or crossmodal adaptation effects typically observed in face perception which used static faces as adaptors (Fox & Barton, 2007; Kovacs et al., 2006). However, there is

some evidence to suggest that *dynamic* faces are likely to be integrated across modalities. For instance, the "McGurk-illusion" which refers to the perceptual fusion of conflicting auditory and visual phonemes, argues for crossmodal audiovisual integration in speech perception (McGurk & MacDonald, 1976). Furthermore, audiovisual integration in voice identity perception is enhanced for synchronized articulating faces as opposed to static images (Schweinberger, Robertson, & Kaufmann, 2007). It was even shown that auditory cortex can be activated in the absence of sound, i.e. by lipreading from silent videos of articulating speakers (Calvert et al., 1997). Therefore, adaptation to silently articulating faces might be considered to activate gender-specific auditory voice representations, hence giving rise to the VGAE. Exp. 3 addressed this possibility.

However, adaptation to videos of silently articulating female or male faces in Exp. 3 did not indicate voice aftereffects suggesting that the VGAE observed in Exp. 1 cannot be explained in terms of adaptation to supramodal female or male speaker representations. Since these experiments had indicated that the VGAE is located within auditory processing, Exp. 4 was designed to test for the possibility that the VGAE reflects low-level adaptation to voice pitch which systematically differed between female (mean F0 = 201 Hz) and male (mean F0 = 108 Hz) voice adaptors in Exp. 1. Another group of participants therefore adapted to sinusoidal tones which were equivalent to F0 of former adaptor voices. Again, no adaptation effects were observed.

Taken together, the four experiments reported in study I provided the first experimental evidence for auditory adaptation to non-linguistic vocal attributes (Exp. 1), a finding that could not be explained by adaptation to gender concepts (Exp. 2), to representations of speaker gender irrespective of adapting modality (Exp. 3) or adaptation to pitch as a low-level stimulus property (Exp. 4). Rather, the VGAE suggests the existence of high-level auditory voice representations selectively tuned to male and female voice quality. While the longevity of face aftereffects is largely unknown, the present results indicate that the recalibration of auditory gender representations following adaptation to voices takes at least a few minutes.

One question arising from the combined results of study I and previous neuroimaging research may be considered as the starting point for study II of the present thesis. If there are neurons specialised in voice gender processing (Lattner et al., 2005; Sokhi et al., 2005) as well as activity decrease in brain areas sensitive to repetitions of individual voices (Belin & Zatorre, 2003), neuronal response suppression should also occur for repetition of the same voice gender. Study II tested this prediction using electrophysiological measures. Another

question following from these considerations is whether there would be a behavioural correlate of voice identity adaptation (Belin & Zatorre, 2003). Therefore, study III set out to test for this possibility and further examined if crossmodal adaptation to faces would induce voice identity aftereffects for speakers with whom participants are familiar.

3.2 In the Ear of the Beholder: Neural Correlates of Adaptation to Voice Gender (Zäske et al., 2009)

Study II was aimed at replicating behavioural results of study I, and at testing whether response suppression in voice gender detectors would be reflected in repetition sensitive event-related potentials (ERPs) of the electroencephalogram (EEG). Modulations of neuronal activity in auditory cortex areas have been demonstrated for repetitions of spectrally complex sounds including animal vocalisations (e.g. Altmann et al., 2008), piano sounds and vowels (e.g. Kuriki, Ohta, & Koyama, 2007) and human voices (Belin & Zatorre, 2003). Specifically, Belin and Zatorre showed that activity in voice selective areas in the right anterior STS can be reduced by repeated stimulation with the voice of the same speaker relative to a silent baseline in fMRI. They concluded that neurons in auditory cortex may be sensitive to individual voice characteristics. Similarly, the VGAE could also be due to response suppression in opponent neuron populations coding female and male voices. Potential neural correlates of adaptation to voice may be the repetition-sensitive N1 and P2 components of the auditory evoked potential (AEP). The N1-P2 complex is a prominent obligatory long-latency response to auditory stimuli. It is measured at fronto-central scalp sites (Vaughan & Ritter, 1970) and thought to be generated in or near primary auditory cortex (Hari, Aittoniemi, Jarvinen, Katila, & Varpula, 1980; Näätänen & Picton, 1987; Scherg & Voncramon, 1985; Wood & Wolpaw, 1982). The negative-going N1 wave peaks at around 100 ms after stimulus while the P2 is a positive deflection with an approximate peak latency of 200 ms (Vaughan & Ritter, 1970).

Repetition-induced attenuation of the N1 increases as a function of similarity between two consecutive simple stimuli which has been attributed to the activity of overlapping neuron populations coding different stimulus features (Näätänen et al., 1988). Similarly, the P2 is modulated by repetitions of emotional prosody (Spreckelmeyer, Kutas, Urbach, Altenmuller, & Muller, 2009) and identity information in voices (Schweinberger, 2001). Being potential neural correlates of the VGAE, study II assessed whether N1 and P2 components would be systematically reduced during test voice presentation when the gender of adaptors and test voices is congruent relative to less congruent conditions.

Study II was conducted in broad analogy to study I (Exp. 1) using the same stimulus material, with adaptation to unfamiliar female and male voices respectively. Speeded gender judgements were performed for test voices varying on a gender continuum between 20/80% and 80/20% proportions in adaptation blocks, while post-adaptation blocks used androgynous voices (50/50%) only. In addition to behavioural responses, I recorded a 32-channel EEG.

As before, I observed a prominent VGAE with more female responses following adaptation to male voices and vice versa with no effects of listener gender. A shift in the category boundary was reflected in PSE of female and male adaptation conditions as well as in reaction times. Participants responded slowest for morph levels approximately corresponding to the respective PSE. This finding is in line with the notion that response uncertainty biases and delays gender decisions as a function of previous adaptation. As expected based on study I (Exp. 1), the VGAE lasted at least a few minutes for androgynous test voices. Overall, the behavioural results of study II replicate the findings of study I (Exp. 1).

In ERPs systematic amplitude attenuations were observed for the N1-P2 complex at frontocentral electrode sites as a function of adaptor-test similarity: Relatively male test voices (20/80% male/female proportion) elicited a significantly smaller central N1 (120-180 ms) when preceded by male rather than female adaptors. Similarly, frontal P2 (210-270 ms) was significantly reduced for female test voices (80/20%) when preceded by female adaptors as opposed to male ones. By contrast, AEPs measured for more ambiguous morph levels of test voice (40/60% and 60/40%) were unaffected by adaptation condition, as were AEPs to androgynous test voices (50/50%) in post-adaptation trials. These findings may reflect the adaptation-induced habituation of contrastive neuron populations coding different aspects of female and male voices. While the N1effect may be due to similarities in pitch, the P2 effect is possibly related to high-level processing of voice spectral information (Chartrand, Peretz, & Belin, 2008; Schröger, 2007). A similar pattern of congruency effects as in N1-P2 was also observed in a later time range (P3, 300-700 ms) but at a more posterior location (Pz). At Pz the effect became significant for both male and female test voices. This may reflect reduced response uncertainty (Johnson, 1984) for test voices towards whose gender, previous adaptation had shifted the internal category boundary.

The lack of significant ERP effects to gender-ambiguous test voices seems in contrast to behavioural aftereffects which were most pronounced for these voices. It is possible, however, that this putative discrepancy is due to spatially overlapping neuron populations coding male and female voices, as indeed has been suggested for male and female faces, or for upright and inverted faces (Jaquet & Rhodes, 2008; Rhodes et al., 2004). Such an account would imply that simultaneous opposite aftereffects could be induced for female and male voices along a different perceptual dimension (e.g. emotional intonation or linguistic contrasts), and that adaptation to androgynous voices should cause only small or negligible aftereffects (Webster & Maclin, 1999). FMRI data suggest that voice gender detectors could be located in the right superior temporal cortex (Lattner et al., 2005) and/or the mesio-parietal

precuneus area at least in male participants (Sokhi et al., 2005). Interestingly, N1 and P2 peak latencies tended to be increased and were significantly increased respectively as test stimuli became more female. This could be tentatively related to higher F0 in female voices and would be expected on the basis of the relationship usually observed between N1(m) latency and frequency in pure tones (Stufflebeam, Poeppel, Rowley, & Roberts, 1998; Woods, Alho, & Algazi, 1993) and complex sounds (Crottaz-Herbette & Ragot, 2000; Ragot & LepaulErcole, 1996).

In future studies ERP measures could be used to determine if other social signals in voices (e.g. identity, age) are also subject to adaptation effects and whether these could be topographically or temporally dissociated from each other. This seems plausible in the face of different though overlapping areas along the STS which are involved in distinct aspects of voice processing (Formisano, De Martino, Bonte, & Goebel, 2008; von Kriegstein & Giraud, 2004). In line with the idea of independent mechanisms underlying the coding of different social signals, Spreckelmeyer and others (2009) showed that the repetition of emotional intonation, but not of voice identity, causes amplitude reductions in P2. Furthermore, voice identity processing can be selectively impaired in cases of congenital and progressive associative phonagnosia with intact recognition of emotional intonation and voice gender (Garrido et al., 2009; Hailstone, Crutch, Vestergaard, Patterson, & Warren, 2010). Accordingly, voice identification and emotion recognition in healthy listeners relies on distinct brain structures (Imaizumi et al., 1997).

3.3 Voice Aftereffects of Adaptation to Speaker Identity (Zäske et al., 2010)

While neural substrates of voice identity processing as likely located in right anterior STS and/or posterior STS (von Kriegstein & Giraud, 2004) are subject to repetition-induced neuronal adaptation at least for unfamiliar voices (Belin & Zatorre, 2003), behavioural correlates of voice identity adaptation are as yet unknown. One objective of study III was therefore to test if adaptation to voices of personally familiar speakers would induce auditory aftereffects in the perception of voice identity in analogy to the VGAE. The second question emerged from a contradiction in the literature concerning the exchange of auditory and visual speaker information. Study I, as well as studies on face adaptation (Fox & Barton, 2007; Kovacs et al., 2006) and speech perception (Green, Kuhl, Meltzoff, & Stevens, 1991), failed to demonstrate crossmodal adaptation or integration effects using unfamiliar faces and voices. By contrast, it has been shown that identity information is integrated across modalities if listeners are familiar with speakers (Schweinberger et al., 2007; Walker, Bruce, & Omalley, 1995). To account for this discrepancy, it has been argued that only for familiar speakers multimodal perceptual person representations exist (Schweinberger et al., 2007). These are thought to be due to ample experience with the simultaneous pairing of facial articulatory movements and vocal signals. This is different from the notion of post-perceptual representations (PINs) which have traditionally been conceived as the locus of face-voice integration (e.g. Burton et al., 1990). By contrast, early audiovisual integration has been indirectly evidenced by crossmodal repetition priming between familiar faces and voices (Ellis et al., 1997; Schweinberger et al., 1997) as well as enhanced voice learning in the presence of dynamic faces (Sheffert & Olson, 2004). Moreover, performance in a voice familiarity task was systematically facilitated or impaired for personally familiar voices (as opposed to unfamiliar ones) when time-synchronized articulating faces of corresponding or non-corresponding speaker identities were presented alongside voices (Robertson & Schweinberger, 2010; Schweinberger et al., 2007).

Study III was thus designed to test if contrastive aftereffects in voice identity perception can be induced by unimodal (Exp. 1) and crossmodal (Exp. 2) adaptation, i.e. by adaptation to voices or faces of personally familiar speakers. Instead of VCV syllables I used two different and ecologically more valid sentence stimuli. Identity-ambiguous test stimuli were created by morphing between voices of two male lecturers personally known to student participants. The test speaker pair was chosen on the basis of a pilot study suggesting that their voices were best discriminable by listeners unfamiliar with them. Adaptor stimuli were voices and videos of silently articulating faces from the same two speakers (A and B) plus a

third equally familiar speaker (C). Dynamic face stimuli rather than static images served as adaptors in Exp. 2 as the former are more likely to be integrated across modalities (Schweinberger et al., 2007). The design of study III was similar to that of a recent study investigating adaptation effects in the perception of facial identity (Fox, Oruc, & Barton, 2008).

Prior to adaptation ambiguous test voices varying between 30/70% and 70/30% of voice identity proportions were predominantly perceived as belonging to speaker B in Exp. 1 and were equally often categorised as originating from both speakers A and B in Exp. 2 in a 2-AFC identification task. This difference may be due to sample effects since procedures and stimuli were identical in both experiments. More importantly, a voice identity aftereffect (VIAE) was seen such that following prolonged exposure to speaker B's voice identity-ambiguous test voices were more often perceived as belonging to speaker A, and vice versa. Adaptation to speaker C induced an intermediate amount of speaker A identifications on the A-B continuum, suggesting that the unimodal VIAE critically depends on shared identity information in adaptor and test voices. As in study I, pitch adaptation can be largely ruled out as being causal for these effects, as all three voices had similar F0. Similarly, speaker differences in specific phonetic sequences or timing could not explain effects as different sentences were used during adaptation and at test and because voice samples were time-standardized with respect to word onsets. The unimodal VIAE which is reminiscent of the face identity aftereffect (FIAE, Leopold et al., 2005), is in line with the idea of identity-specific voice representations in long-term memory which habituate with repeated stimulation.

Interestingly and in contrast to study I (Exp. 3), an analogous pattern of aftereffects on familiar voice identity processing was now also seen following face adaptation (Exp. 2), i.e. in the absence of acoustic stimulation. Aftereffects were reflected in significantly more speaker A classifications following B vs. A adaptation. This finding argues for the existence of relatively flexible multimodal perceptual representations of familiar speakers. As a qualification, the crossmodal VIAE was not only smaller in magnitude than unimodal aftereffects but also less persistent: whereas, similar to the VGAE, biased voice identity perception was still evident a few minutes after adaptation in Exp. 1, effects had completely dissipated in Exp. 2.

These differences in magnitude and longevity of unimodal and crossmodal effects are in agreement with repetition priming data (Ellis et al., 1997; Schweinberger et al., 1997) and may be indicative of at least two mechanisms of identity adaptation: The first may be related

to auditory coding of voice characteristics as possibly mediated by right anterior STS areas. These areas are sensitive to speaker repetitions irrespective of speech content (Belin & Zatorre, 2003) and are preferentially activated during voice recognition rather than semantic content recognition irrespective of speaker familiarity (von Kriegstein & Giraud, 2004). The second mechanism, by contrast, could be related to multimodal coding of familiar speaker identity. Such multimodal coding could be implemented either via direct connections between face- and voice-sensitive areas (von Kriegstein & Giraud, 2006; von Kriegstein et al., 2005), or via a separate multimodal pool of neurons. Further research will be needed to assess these alternative hypotheses. Multimodal areas could be located in posterior STS/MTG as suggested by studies investigating face-voice integration in emotional expression analysis (Kreifelts, Ethofer, Grodd, Erb, & Wildgruber, 2007) and multimodal recognition of familiar environmental sound sources (Lewis et al., 2004).

4 General Discussion

The three studies reported here provide novel evidence for high-level auditory aftereffects in the perception of *non-linguistic* vocal signals, i.e. voice gender and speaker identity. This implies that recent experience with a certain group of voices (i.e. male or female) or with individual voices (or faces) can temporarily bias auditory perception of speaker attributes in a behaviourally relevant way. Findings of contrastive aftereffects in the perception of voice gender and identity complement and extend analogous reports from face perception research (e.g. Leopold et al., 2001; Webster et al., 2004) and suggest that adaptation routinely influences the perception of social information from both faces and voices. Voice aftereffects further add to auditory adaptation effects for *linguistic* properties of speech, which have been known for some time (e.g. Eimas & Corbit, 1973; Morse et al., 1976). Moreover, the present findings corroborate the idea that adaptation is a ubiquitous mechanism in human perception that applies both to the processing of simple *and* complex stimulus features. Finally, the present results point to the existence of neurons selectively tuned to particular aspects of voice quality.

Adaptation to vocal age – preliminary data

One question following from this line of argument is whether other social signals in voices such as age, are subject to analogous coding principles. Preliminary and as yet unpublished data from 80 student participants suggest that adaptation to young and elderly voices biases age judgements for various morph levels of test voice away from the adapting age (Zäske & Schweinberger, 2010, cf. Appendix 1). Thus, test voices with interpolated ages of approximately 30, 40, 50, or 60 yrs are perceived as older with previous adaptation to young voices (20 yrs) relative to a condition of adaptation to old voices (70 yrs). This finding complements the recent discovery of facial age aftereffects (FAAE, Schweinberger et al., 2009) and argues for the existence of opponent voice age detectors tuned to young (20 yrs) and old (70yrs) voices respectively.

Similar to other auditory aftereffects reported here, vocal age aftereffects (VAAE) remained significant for at least a few minutes. At the same time, however, the magnitude of these effects was significantly reduced in post-adaptation blocks relative to adaptation blocks when speakers and listeners had different genders as opposed to the same gender. A tentative interpretation for this gender-congruency effect in the longevity of the VAAE is that a fast recalibration of vocal age perception for other-gender individuals may facilitate mate selection. In line with honest signal hypothesis (Johnstone, 1995; Zahavi, 1977), it has been

suggested that voices not only reliably indicate speaker age (Bruckert et al., 2006) but also signal mate quality and reproductive success in humans (Apicella, Feinberg, & Marlowe, 2007; Hughes, Dispenza, & Gallup, 2004). Furthermore, the same acoustic cues appear to affect the assessment of speaker attractiveness and age (Feinberg, Jones, Little, Burt, & Perrett, 2005). Specifically, this group increased pitch and decreased apparent vocal tract length of male speakers' voices (aged 20-22 yrs) by altering F0 and formant frequencies. Following these manipulations, women rated speakers as smaller, younger, and as less attractive than speakers whose voices that had been re-edited to match the originals. Feinberg and colleagues interpreted their finding as reflecting female preference for male voices which signal full sexual maturity. It is well known that as age further advances, reproductive fitness decreases in adults. Therefore, in order to allow for a more reliable assessment of mate quality following voice adaptation, it may be crucial for listeners to instantly regain the ability of estimating age from voices of other-gender individuals.

A number of predictions follow from this line of thought. First, congruency-effects of speaker and listener gender should modulate the persistence of other voice aftereffects, provided the features inducing these aftereffects are relevant for mate selection or the perception of attractiveness. Second, given that the correct perception of gender from voices is equally important for male and female listeners, the VGAE should be subject to an equal rate of decay for both listener genders. In line with this reasoning, studies I and II of the present thesis did not suggest any effects of listener gender on the magnitude of the VGAE, i.e. neither in adaptation nor in post-adaptation phases. Third, gender-congruency effects of speaker and listener should also be evident in other stimuli that are implicated in mate evaluation, for instance faces (Saxton, DeBruine, Jones, Little, & Roberts, 2009). Unfortunately, the only study investigating age aftereffects in face perception so far (Schweinberger et al., 2009) did not test for the longevity of the FAAE.

A prototype account of voice aftereffects

In face perception research, contrastive aftereffects have been implemented to probe norm-based coding of face identity (Leopold et al., 2001; Rhodes & Jeffery, 2006). These studies support the notion that faces are represented as points (or vectors) in a multidimensional space. Within that space, the position of each face is defined by its deviation from a central average face, i.e. the prototype (Valentine, 1991). The coordinates of a given face are defined by its values on an unspecified number of feature dimensions. According to Valentine, these dimensions or "axes" signify single features such as hair colour

or size of the nose. By contrast, in Leopold's et al. (2001) terminology, axes denote so-called identity trajectories each of which connects two faces located at opposite ends of the space. Faces occupying homologous points of the same identity trajectory (i.e. faces and "anti-faces"), differ from the prototype to the same degree but in opposite directions with respect to the features that distinguish them from the prototype. In their seminal study, Leopold and colleagues (2001) generated a face space whose "prototype" was a morph of 200 different face images. They demonstrated that perception of this average face was biased towards a particular known person (e.g. "Jim") after adaptation to an unfamiliar face ("anti-Jim") with opposite features. This suggests that adaptation can shift the perceptual boundary of face identity, i.e. the prototype, towards the adapting face along the respective identity trajectory. A parsimonious account for the FIAE is that each identity trajectory is coded by only two cell types which are each maximally responsive to a different endpoint of a trajectory. The coding of all other faces lying on that trajectory would be achieved by different activation levels in these cells (Tsao & Freiwald, 2006).

Prototype models have also been advanced to describe the mental representation of voice identity, both for unfamiliar speakers (Papcun, Kreiman, & Davis, 1989) and familiar speakers (Lavner et al., 2001). However, testing of these models using auditory adaptation paradigms has been impeded by technical limitations of high-quality voice averaging across a large number of exemplars. Study III (Exp. 1) of the present thesis demonstrated VIAEs on an identity continuum between only two familiar voices. Naturally, test voice averages between two male speakers provide only a poor approximation of a population average. Therefore, the present VIAE may point to the operation of voice detectors selectively tuned to different speaker identities without, however, making a strong argument for norm-based coding.

Interestingly, one group recently generated male and female voice averages of the syllable "had" as derived from 32 speakers of each gender and tested the perception of vocal attractiveness of (Bruckert et al., 2010). Attractiveness ratings for various voice composites increased with the number of voices contributing to the composites. Bruckert and colleagues interpreted their findings in terms of a great similarity between highly-attractive voices and a voice prototype. They further suggested that voices may be encoded relative to this prototype. In support of this claim, voice identity aftereffects of adaptation to 'anti-voices' (Latinus & Belin, 2010) have recently been reported in broad analogy to Leopold et al. (2001). Taking Tsao's and Freiwald's (2006) account of the FIAE as a basis, Latinus' and Belin's findings could be tentatively interpreted in terms of an adaptation-induced activity imbalance in neurons contrastively coding the endpoints of voice identity trajectories.

At present, a prototype account in analogy to Valentine's model (1991) would seem the most useful framework for understanding gender and age aftereffects in voice perception. However, before applying the "voice space heuristic" to the present results, I will turn to a longstanding controversy concerning the question of what the axes in Valentine' model exactly represent. Valentine (1991, p. 168) vaguely defined axes as dimensions "[…] that could be used to discriminate faces". This may imply separate dimensions for simple stimulus features but could also refer to more complex features such as gender and age (Johnston, Kanazawa, Kato, & Oda, 1997). Clearly, social signals are complex features that are confounded with (or determined by) more or less typical low-level stimulus configurations in both voices (cf. chapter 1) and faces. Therefore, instead of treating simple and complex features as equivalent coexisting axes, it seems more plausible to assume simple feature dimensions, which determine the location of voice populations sharing common features. These voice clusters could for instance represent various age groups or female and male voices respectively. Note, that a similar suggestion has been made by Johnston et al. (1997), apart from the fact that these authors proposed separate dimensions for face gender and age.

In analogy to Valentine's (1991) proposal, a voice prototype would reflect the central tendency of all voices ever heard, whether or not listeners were able to identify them. Depending on the nature of voices encountered throughout lifetime, the prototype could "sound" like an androgynous middle-aged voice. At least this would be expected if the listener had experienced an equal amount of male and female voices as well as young and old voices. Forasmuch as gender and age categories differ on low-level feature dimensions (e.g. F0, jitter, breathiness, speaking rate, etc.), voices of opponent qualities should be found in more or less different "corners" of the space. Adaptation to voices from one of these categories should suppress activity in voice detectors that share stimulus attributes of the adaptor. Accordingly, the VGAE in studies I and II could be explained in terms of a diffuse activity decrease in female or male voice representations that have been triggered by similar adaptor voices. In this way, gender aftereffects can even be induced by unfamiliar voices which do not have an individual representation in the voice space or which may be about to obtain it in the course of adaptation.

Borrowing from Leopold's et al. (2001) concept of identity trajectories, it could be assumed that clusters of male and female voices are connected by gender trajectories spanning across the voice space. In line with the idea of widespread response suppression in neurons coding similar voices (e.g. female voices), the VGAE survived changes of speaker identities during adaptation (study 1, Exp. 1) as well as between adaptation and test (studies I and II).

With respect to age perception, preliminary findings of VAAEs could be interpreted in analogy to gender aftereffects. However, the differential persistence of the VAAE related to the gender of the speaker and listener would be difficult to explain within this working model.

Note, that all voice aftereffects reported here, were measured following adaptation to more than just one social signal at a time. For instance, male and female adaptors in studies I and II were young adult speakers. Similarly, personally familiar voice adaptors in study III not only represented particular identities, but also transmitted gender and age information. Since voices contain various speaker attributes, adaptation to one feature usually implies adaptation to all signals common to adaptors. Which aftereffect we observe in a given adaptation paradigm, depends on two main factors: (1) the task at test (e.g. speaker identification, gender categorisation) and (2) the presence of at least two to-be-compared adaptation conditions, which manipulate social information along the test dimension. (Instead of the second adaptation condition or in addition to it, a pre-adaptation baseline could serve as a reference measure of adaptation). With respect to the prototype account, simultaneous adaptation to various social signals would entail complex patterns of inhibition in the "multidimensional voice space".

While the notion of multidimensional face or voice spaces may be compatible with unimodal adaptation effects in person perception, there is no framework at present, which could fully account for the crossmodal voice aftereffects observed in study III (Exp. 2). However, while prototype approaches (sensu Leopold et al., 2001; Valentine, 1991) serve to model the contrastive nature of adaptation effects, neurocognitive models (e.g. Belin et al., 2004; Campanella & Belin, 2007) incorporate the idea of crossmodal exchange via links between "unimodal" face and voice modules. One way of merging these ideas would be the assumption of two spaces, one for faces and one for voices. If points represent faces and voices of speakers with whom there exists bimodal communication experience, corresponding points of speaker identity would be somehow connected between spaces. Alternatively, there might be a single "audiovisual" space with points representing either unimodal or multimodal person representations, depending on whether a listener is familiar with a speaker's voice, or face, or both. Based on these heuristics, crossmodal identity aftereffects in the manner of Leopold's et al. findings (2001) would be imaginable.

Top-down effects in voice adaptation

In objection to the claim that adaptation effects are stimulus-driven, it could be argued that these effects are merely a result of top-down processing. For instance, anchoring effects

or mental imagery could be responsible for the findings reported here. However, if voice aftereffects were due to expectation of change following repeated stimulation with one stimulus type, there should have been crossmodal effects of adaptation to names or faces on gender perception in study I (Exp. 2 and 3). In order to test the possible involvement of expectations in the genesis of VIAEs (study III), an unpublished control experiment was conducted on 12 participants who adapted to names of personally familiar speakers. In line with crossmodal null results in study I, adaptation to names did not elicit any voice aftereffects in that control condition (cf. Appendix 2). At the same time, these preliminary data indirectly strengthen the argument that the crossmodal VIAE as induced by faces is due to the operation of multimodal perceptual speaker representations. These representations likely exist for faces and voices (but not for names), as these stimuli are usually encountered in close temporal coincidence.

At present, however, one cannot exclude the possibility that auditory imagery was involved in crossmodal voice aftereffects in study III (Exp. 2). It is feasible to assume that the rare sight of a talking and yet mute familiar face causes mental imagery of the respective voice. Possibly related to this account, two studies found auditory cortex activation (Calvert et al., 1997) and suppressed N1(m) responses from auditory cortex neurons (Jääskeläinen et al., 2004) in response to silent visual speech. It is further known that patients experiencing voice hallucinations exhibit activity in auditory sensory cortices (for review see Allen, Laroi, McGuire, & Aleman, 2008). Similarly, it has been suggested that voice-induced activation of face areas in healthy listeners may be accompanied by mental imagery of a speaker's face (von Kriegstein et al., 2005).

To the best of my knowledge, there are no studies investigating crossmodal effects of adaptation to imagined voices or faces. However, face perception research offers some contradicting findings on the effects of unimodal adaptation to imagined familiar (famous) faces (DeBruine, Welling, Jones, & Little, 2010; Ganis & Schendan, 2008; Ryu, Borrmann, & Chaudhuri, 2008). One study provides evidence in favour of contrastive identity aftereffects following either imagined or real face adaptors (Ryu et al., 2008) suggesting that common neural networks mediate perception and volitional imagery of faces. At variance with these findings, contrastive aftereffects in gender perception were only observed for real face adaptors, whereas imagined faces produced the opposite pattern of results, i.e. facilitatory priming (DeBruine et al., 2010). In line with these latter findings, EEG research had suggested that adaptation to visualized or real face images differentially affects early face-sensitive ERPs in response to famous test faces (Ganis & Schendan, 2008). While in that

study, amplitudes of the N170 and the vertex positive potential (VPP) were suppressed when test faces had been preceded by photos of the same celebrities, N170/VPP amplitudes were enhanced following imagery of these faces. This was taken to suggest that neuronal populations mediating early face processing are sensitive to bottom-up input from perceived adaptors as well as to top-down input from visualized face adaptors.

Whether or not face-induced voice imagery elicited or affected crossmodal aftereffects in study III could be probed by having participants imagine familiar speakers' voices during adaptation. If the crossmodal VIAE was due to auditory imagery, perception of identity-ambiguous test voices should be subject to contrastive aftereffects (rather than priming). Furthermore, if mental imagery was the sole source of the crossmodal VIAE, aftereffects of adaptation to imagined voices should be equal in magnitude to those elicited by faces.

5 Outlook

Measures of voice perception are affected by numerous factors such as the stimulus dimension that is being tested (Collins, 2000; Fecteau et al., 2004), the stimulus variety and/or duration (Cook & Wilding, 1997; Pollack, Pickett, & Sumby, 1954; Schweinberger, Herholz, & Sommer, 1997) or the kind of stimuli used (Lass et al., 1976). Further examples include presentation conditions and specific task demands during learning and recall, retention intervals (Clifford, 1980), and the listeners' expertise with voices (Beauchemin et al., 2006; Kreiman, Gerratt, & Precoda, 1990) or with other classes of auditory objects (Chartrand & Belin, 2006; Chartrand, Filion-Bilodeau, & Belin, 2007). Last but not least, performances of voice perception may depend on speaker characteristics, both on an individual level (Papcun et al., 1989) and on a group level (Fecteau et al., 2004). Together with many others, these factors could be subsumed under three broad categories, these being characteristics of (1) the speaker, (2) the listener and (3) external conditions affecting different processing stages of voice perception and memory. Adding to these factors, the present studies provide converging evidence that perception of gender and identity information in voices is significantly affected by the kind of voices (or faces) recently encountered. Some unresolved questions related to the present research (and related to voice perception research in general) shall be briefly addressed in the following.

Are voice representations for different social signals independent of each other?

Traditionally, processing routes for different types of social information in faces such as identity, gender and expression, have been thought to be independent of each other (Bruce & Young, 1986). Meanwhile, this claim has been challenged by a number of studies reporting processing dependencies between for instance, facial expression and identity (Fox & Barton, 2007; Schweinberger & Soukup, 1998) or gender and age (Barrett & O'Toole, 2009; Kloth, Wiese, & Schweinberger, 2010; Schweinberger et al., 2009). With respect to the voice domain, there is some evidence from clinical case studies suggesting that voice identity is perceived independent of voice gender and emotional prosody (Garrido et al., 2009; Hailstone et al., 2010). However, based on the assumption that many social signals are available simultaneously from voices and at least partially rely on the same acoustic cues, it seems plausible that their processing is not entirely independent. In line with this reasoning, Burton and Bonner (2004) found that healthy listeners classified the gender of voices faster when they were familiar with them. These findings indicate that voice gender perception is affected by voice familiarity although gender information is equally present in both types of voices.

Recently, variants of the classic adaptation paradigm have proven fruitful tools to investigate whether two feature dimensions (e.g. face gender and age) or two face groups (e.g. male and female faces) are subject to common or selective neural coding mechanisms (e.g. Jaquet, Rhodes, & Hayward, 2007; Little, DeBruine, & Jones, 2005; Rhodes et al., 2004; Schweinberger et al., 2009). For instance, the significant though reduced transfer of facial age aftereffects across different adaptor and test genders points to both gender-independent and gender-contingent coding of face age (Schweinberger et al., 2009).

With respect to the present research, it was certainly sensible to initially focus on one social signal at a time. However, it may be a promising next step to utilize adaptation paradigms for investigating coding dependencies between different vocal signals. Based on the "multidimensional voice space" heuristic, such coding dependencies would be predicted for instance, for voice gender and age. Voice clusters denoting different social categories should partly overlap in the voice space due to shared low-level feature values. For instance, voices with similar pitch (i.e. F0 values) which would be represented close to each other may nevertheless be representatives of entirely different age and gender groups, i.e. boys or adult women. Based on this reasoning, it should be possible to observe at least partial transfer of the VGAE across different adaptor and test ages as has been suggested for faces (Barrett & O'Toole, 2009). Conversely, the VAAE may survive adaptation-to-test changes of voice gender in analogy to the FAAE (Schweinberger et al., 2009). Category-contingent coding, on the other hand, should be evident in simultaneous opposite voice aftereffects in analogy to face aftereffects (e.g. Jaquet & Rhodes, 2008; Rhodes et al., 2004).

The role of listener attributes in voice perception

Another promising field of research may be listener characteristics in voice perception. Some studies (Lattner et al., 2005; Owren et al., 2007) including the present ones (I and II) failed to find behavioural and/or neurophysiological effects of listener gender in tasks of voice gender perception. However, listener gender may affect the longevity of VAAEs (chapter 4). Furthermore, one group (Schirmer, Striano, & Friederici, 2005) reported effects of listener gender on the preattentive detection of change in vocal emotion. In an oddball paradigm, female participants showed a larger MMN than male participants in response to rare syllables spoken in emotional tone within a stream of neutrally spoken syllables. This was interpreted in terms of an additional recruitment of processing resources in women when voices deviants had an emotional valence. Furthermore, there is evidence from low-level audition according

to which women are more susceptible to loudness adaptation than men (D'Alessandro & Norwich, 2009).

In sum, these findings tentatively suggest that the emergence of listener gender effects may depend on the feature dimension relevant to the task, the level of processing required and the dependent measures used. Other, though rare examples for listener variables that have been investigated by voice researchers include age (Kausler & Puckett, 1981; Linville & Korabic, 1986; Yonan & Sommers, 2000), auditory expertise (Chartrand & Belin, 2006; Chartrand et al., 2007), language abilities and ethnicity (Perrachione, Chiao, & Wong, 2010; Perrachione, Pierrehumbert, & Wong, 2009), visual abilities (Gougoux et al., 2009) as well as neurological conditions (e.g. Garrido et al., 2009; Hailstone et al., 2010; Mark & Stegman, 1991; Vanlancker et al., 1988; Vanlancker, Kreiman, & Cummings, 1989). However, still very little is known about effects of listener characteristics on speaker perception and memory.

Should we strictly analogize face and voice perception?

Models of person perception propose that faces and voices are subject to analogous processing and storage mechanisms (e.g. Belin et al., 2004; Neuner & Schweinberger, 2000). However, people are generally better at recognising and identifying others from their faces than from their voices (Hanley, Smith, & Hadfield, 1998; Hanley & Turner, 2000) and are also more confident in doing so (Olsson, Juslin, & Winman, 1998). Moreover, Burton and Bonner (2004) highlighted another possible discrepancy in the processing of voices and faces by demonstrating that the speed of voice gender classifications is modulated by voice familiarity. By contrast, gender decisions for faces appear to be independent of familiarity (Bruce, Ellis, Gibling, & Young, 1987). Given these differences, it would be worthwhile to compare the behaviour of faces and voices directly under various experimental conditions and to reconsider current models of person perception.

Summary

Apart from being carriers of speech, human voices contain a wealth of social signals, for instance about a speaker's gender, identity, or age, to name but a few. The present thesis is concerned with the way adaptation modifies the perception of gender and identity information from voices. Adaptation is a mechanism by which neural responses decrease after continuous or repetitive stimulation. It is revealed by transient perceptual aftereffects indicating contrastive coding of simple and complex stimulus properties. The three studies reported here investigate unimodal and crossmodal auditory voice aftereffects of adaptation to unfamiliar and personally familiar speakers.

Specifically, study I (Exp. 1) shows that adaptation to unfamiliar voices of female or male speakers biases the perception of voice gender away from the adapting gender: test voices, as created by auditory morphing between male and female voices, are perceived as more male following adaptation to female voices and vice versa. The voice gender aftereffect (VGAE) survived at least a few minutes and suggests the existence of voice detectors tuned to female and male voice quality. The absence of voice aftereffects following adaptation to names (Exp. 2), faces (Exp. 3), or sinusoidal tones matched to F0 of adaptor voices (Exp. 4) further suggests that the VGAE is due to habituation of high-level auditory representations.

Study II replicates behavioural findings of study I (Exp. 1) and further supports the notion of processing selectivity for female and male voices by providing electrophysiological evidence. Systematic adaptation-induced amplitude reductions in AEPs (N1, P2, and P3) were observed in response to otherwise identical test voices when test voices and adaptors had the same gender as opposed to different genders. This suggests that contrastive coding of voice gender is implemented by auditory cortex neurons and takes place within the first few hundred milliseconds from voice onset.

Similar to the VGAE, auditory aftereffects of adaptation to voices or faces of personally familiar speakers caused contrastive aftereffects in listeners' perception of voice identity (study III). Unimodal voice-to-voice aftereffects (Exp. 1) were more pronounced and more persistent than crossmodal face-to-voice aftereffects (Exp. 2) pointing to at least two perceptual mechanisms of voice identity adaptation: one related to auditory coding of voice characteristics and one related to multimodal coding of speaker identity.

These results complement findings in face perception (z.B. Leopold et al., 2001; Webster et al., 2004) and suggest that adaptation is a ubiquitous mechanism that routinely influences the perception of non-linguistic social information from both faces and voices.

Zusammenfassung

Stimmen sind nicht nur Träger von Sprache, sondern enthalten auch eine Vielzahl von sozialen Signalen, die beispielsweise über das Geschlecht, die Identität oder das Alter eines Sprechers Auskunft geben. Die vorliegende Dissertationsschrift beschäftigt sich mit Adaptationseffekten in der Wahrnehmung von Geschlechts- und Identitätsinformation in menschlichen Stimmen. Adaptation ist ein Mechanismus, bei dem die neuronale Aktivierung infolge von kontinuierlicher oder wiederholter Stimulation abnimmt. Auf Verhaltensebene spiegelt sich Adaptation in vorübergehenden Wahrnehmungsverzerrungen wider. Diese perzeptuellen Nacheffekte verweisen auf die kontrastive neuronale Kodierung von einfachen und komplexen Reizmerkmalen durch sog. Opponenteneinheiten. Die drei berichteten Studien untersuchen intra- und intermodale Stimmennacheffekte nach Adaptation an unbekannte und persönlich bekannte Sprecher.

Studie I (Exp. 1) zeigt, dass Adaptation an Stimmen unbekannter Sprecher und Sprecherinnen die Wahrnehmung von darauffolgenden Teststimmen in Richtung des nichtadaptierten Geschlechts verzerrt. Nach wiederholter Darbietung von Männerstimmen, wurden die entlang eines Geschlechtskontinuums variierenden Teststimmen als weiblicher wahrgenommen und umgekehrt. Der „voice gender aftereffect“ (VGAE) war auch einige Minuten nach Adaptation noch messbar und deutet auf die Existenz von Stimmendetektoren hin, die selektiv weibliche und männliche Stimmqualität kodieren. Das Ausbleiben von Stimmennacheffekten bei Adaptation an Namen (Exp. 2), Gesichter (Exp. 3) oder an Sinustöne mit „weiblicher“ und „männlicher“ Grundfrequenz spricht dafür, dass der VGAE auf die Habituierung von Stimmenrepräsentationen auf höheren Stufen der auditiven Informationsverarbeitung zurückgeht.

Studie II konnte die Verhaltensbefunde von Studie I (Exp. 1) replizieren und liefert elektrophysiologische Evidenz für die Annahme von antagonistischen Stimmendetektoren, die auf die Verarbeitung von Geschlechtsinformation in Stimmen spezialisiert sind. Als Antwort auf Testimmen wurden systematische adaptationsbedingte Amplitudenreduktionen in akustisch evozierten Potentialen (N1, P2 und P3) des Elektroenzephalogramms gemessen, wenn sich das Geschlecht von Test- und Adaptorstimmen ähnelten, im Vergleich zu Bedingungen, in denen es sich unterschied. Dies legt nahe, dass die Kodierung des Stimmengeschlechts von Neuronen im auditorischen Kortex realisiert wird und innerhalb der ersten paar hundert Millisekunden nach Beginn der Stimmendarbietung stattfindet.

Ähnlich dem VGAE, lassen sich kontrastive auditive Nacheffekte auch für die Wahrnehmung von Sprecheridentität induzieren. In Studie III adaptierten Probanden entweder

an Stimmen (Exp. 1) oder an lautlos artikulierende Gesichter (Exp. 2) persönlich bekannter Sprecher. Nach Adaptation an die Stimme oder das Gesicht von Sprecher A wurden identitätsambige Teststimmen-Morphs aus Sprecher A und B eher Sprecher B zugeordnet und umgekehrt. Hingegen rief Adaptation an einen unrelatierten Sprecher C keinen Nacheffekt hervor. Dabei war der intramodale „voice identity aftereffect“ (VIAE) ausgeprägter und langlebiger als der intermodale VIAE, was auf die Beteiligung von mindestens zwei Mechanismen hindeutet. Diese könnten jeweils mit der auditiven Kodierung von Stimmenmerkmalen bzw. mit der multimodalen Kodierung von Sprecheridentität im Zusammenhang stehen.

Insgesamt passen die Ergebnisse dieser Studien zu ähnlichen Befunden aus der Gesichterforschung (e.g. Leopold et al., 2001; Webster et al., 2004) und legen nahe, dass Adaptation ein universeller Mechanismus in der Wahrnehmung von nichtlinguistischer sozialer Information in Gesichtern und Stimmen ist.

References

Allen, P., Laroi, F., McGuire, P. K., & Aleman, A. (2008). The hallucinating brain: A review of structural and functional neuroimaging studies of hallucinations. *Neuroscience and Biobehavioral Reviews, 32,* 175-191.

Altmann, C. F., Nakata, H., Noguchi, Y., Inui, K., Hoshiyama, M., Kaneoke, Y. et al. (2008). Temporal dynamics of adaptation to natural sounds in the human auditory cortex. *Cerebral Cortex, 18,* 1350-1360.

Anstis, S., Verstraten, F. A. J., & Mather, G. (1998). The motion aftereffect. *Trends in Cognitive Sciences, 2,* 111-117.

Apicella, C. L., Feinberg, D. R., & Marlowe, F. W. (2007). Voice pitch predicts reproductive success in male hunter-gatherers. *Biology Letters, 3,* 682-684.

Barrett, S. E. & O'Toole, A. J. (2009). Face adaptation to gender: Does adaptation transfer across age categories? *Visual Cognition, 17,* 700-715.

Baumann, O. & Belin, P. (2010). Perceptual scaling of voice identity: common dimensions for different vowels and speakers. *Psychological Research, 74,* 110-120.

Beauchemin, M., De Beaumont, L., Vannasing, P., Turcotte, A., Arcand, C., Belin, P. et al. (2006). Electrophysiological markers of voice familiarity. *European Journal of Neuroscience, 23,* 3081-3086.

Bee, M. A. & Gerhardt, H. C. (2002). Individual voice recognition in a territorial frog (Rana catesbeiana). *Proceedings of the Royal Society of London Series B-Biological Sciences, 269,* 1443-1448.

Belin, P. (2006). Voice processing in human and non-human primates. *Philosophical Transactions of the Royal Society B-Biological Sciences, 361,* 2091-2107.

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

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

Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cognitive Brain Research, 13,* 17-26.

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

Birkett, P. B., Hunter, M. D., Parks, R. W., Farrow, T. F., Lowe, H., Wilkinson, L. D. et al. (2007). Voice familiarity engages auditory cortex. *Neuroreport, 18,* 1375-1378.

Bricker, P. D. & Pruzansky, S. (1966). Effects of Stimulus Content and Duration on Talker Identification. *Journal of the Acoustical Society of America, 40,* 1441-&.

Bruce, V., Ellis, H., Gibling, F., & Young, A. (1987). Parallel Processing of the Sex and Familiarity of Faces. *Canadian Journal of Psychology-Revue Canadienne de Psychologie, 41,* 510-520.

Bruce, V. & Young, A. (1986). Understanding Face Recognition. *British Journal of Psychology, 77,* 305-327.

Bruckert, L., Bestelmeyer, P., Latinus, M., Rouger, J., Charest, I., Rousselet, G. A. et al. (2010). Vocal Attractiveness Increases by Averaging. *Current Biology, 20,* 116-120.

Bruckert, L., Lienard, J. S., Lacroix, A., Kreutzer, M., & Leboucher, G. (2006). Women use voice parameters to assess men's characteristics. *Proceedings of the Royal Society B-Biological Sciences, 273,* 83-89.

Budd, T. W., Barry, R. J., Gordon, E., Rennie, C., & Michie, P. T. (1998). Decrement of the N1 auditory event-related potential with stimulus repetition: habituation vs refractoriness. *International Journal of Psychophysiology, 31,* 51-68.

Burton, A. M. & Bonner, L. (2004). Familiarity influences judgments of sex: The case of voice recognition. *Perception, 33,* 747-752.

Burton, A. M., Bruce, V., & Johnston, R. A. (1990). Understanding Face Recognition with An Interactive Activation Model. *British Journal of Psychology, 81,* 361-380.

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C. R., McGuire, P. K. et al. (1997). Activation of auditory cortex during silent lipreading. *Science, 276,* 593-596.

Campanella, S. & Belin, P. (2007). Integrating face and voice in person perception. *Trends in Cognitive Sciences, 11,* 535-543.

Charest, I., Pernet, C. R., Rousselet, G. A., Quinones, I., Latinus, M., Fillion-Bilodeau, S. et al. (2009). Electrophysiological evidence for an early processing of human voices. *Bmc Neuroscience, 10*.

Chartrand, J. P. & Belin, P. (2006). Superior voice timbre processing in musicians. *Neuroscience Letters, 405,* 164-167.

Chartrand, J. P., Filion-Bilodeau, S., & Belin, P. (2007). Brain response to birdsongs in bird experts. *Neuroreport, 18,* 335-340.

Chartrand, J. P., Peretz, I., & Belin, P. (2008). Auditory recognition expertise and domain specificity. *Brain Research, 1220,* 191-198.

Clifford, B. R. (1980). Voice identification by human listeners: on earwitness reliability. *Law and Human Behavior, 4,* 373-394.

Clifford, C. W. G. & Rhodes, G. (2005). *Fitting the mind to the world: adaptation after-effects in high-level vision*. New York: Oxford University Press.

Collins, S. A. (2000). Men's voices and women's choices. *Animal Behaviour, 60,* 773-780.

Cook, S. & Wilding, J. (1997). Earwitness testimony: Never mind the variety, hear the length. *Applied Cognitive Psychology, 11,* 95-111.

Crottaz-Herbette, S. & Ragot, R. (2000). Perception of complex sounds: N1 latency codes pitch and topography codes spectra. *Clinical Neurophysiology, 111,* 1759-1766.

D'Alessandro, L. M. & Norwich, K. H. (2009). Loudness adaptation measured by the simultaneous dichotic loudness balance technique differs between genders. *Hearing Research, 247,* 122-127.

Dange, A., Warm, J. S., Weiler, E. M., & Dember, W. N. (1993). Loudness Adaptation - Resolution of A Psychophysical Enigma. *Journal of General Psychology, 120,* 217-243.

DeBruine, L. M., Welling, L. L. M., Jones, B. C., & Little, A. C. (2010). Opposite effects of visual versus imagined presentation of faces on subsequent sex perception. *Visual Cognition, 18,* 816-828.

Decasper, A. J. & Fifer, W. P. (1980). Of Human Bonding - Newborns Prefer Their Mothers Voices. *Science, 208,* 1174-1176.

Dong, C. J., Swindale, N. V., & Cynader, M. S. (1999). A contingent aftereffect in the auditory system. *Nature Neuroscience, 2,* 863-865.

Eimas, P. D. & Corbit, J. D. (1973). Selective Adaptation of Linguistic Feature Detectors. *Cognitive Psychology, 4,* 99-109.

Ellis, H. D., Jones, D. M., & Mosdell, N. (1997). Intra- and inter-modal repetition priming of familiar faces and voices. *British Journal of Psychology, 88,* 143-156.

Fecteau, S., Armony, J. L., Joanette, Y., & Belin, P. (2004). Priming of non-speech vocalizations in male adults: The influence of the speaker's gender. *Brain and Cognition, 55,* 300-302.

Feinberg, D. R., Jones, B. C., Little, A. C., Burt, D. M., & Perrett, D. I. (2005). Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices. *Animal Behaviour, 69,* 561-568.

Fitch, W. T. (2000). The evolution of speech: a comparative review. *Trends in Cognitive Sciences, 4,* 258-267.

Formisano, E., De Martino, F., Bonte, M., & Goebel, R. (2008). "Who" Is Saying "What"? Brain-Based Decoding of Human Voice and Speech. *Science, 322,* 970-973.

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

Fox, C. J., Oruc, I., & Barton, J. J. S. (2008). It doesn't matter how you feel. The facial identity aftereffect is invariant to changes in facial expression. *Journal of Vision, 8*.

Ganis, G. & Schendan, H. E. (2008). Visual mental imagery and perception produce opposite adaptation effects on early brain potentials. *Neuroimage, 42,* 1714-1727.

Garrido, L., Eisner, F., McGettigan, C., Stewart, L., Sauter, D., Hanley, J. R. et al. (2009). Developmental phonagnosia: A selective deficit of vocal identity recognition. *Neuropsychologia, 47,* 123-131.

Geiselman, R. E. & Bellezza, F. S. (1976). Long-Term-Memory for Speakers Voice and Source Location. *Memory & Cognition, 4,* 483-489.

Geiselman, R. E. & Bellezza, F. S. (1977). Incidental Retention of Speakers Voice. *Memory & Cognition, 5,* 658-665.

Gelfer, M. P. & Mikos, V. A. (2005). The relative contributions of speaking fundamental frequency and formant frequencies to gender identification based on isolated vowels. *Journal of Voice, 19,* 544-554.

Goggin, J. P., Thompson, C. P., Strube, G., & Simental, L. R. (1991). The Role of Language Familiarity in Voice Identification. *Memory & Cognition, 19,* 448-458.

Gougoux, F., Belin, P., Voss, P., Lepore, F., Lassonde, M., & Zatorre, R. J. (2009). Voice perception in blind persons: A functional magnetic resonance imaging study. *Neuropsychologia, 47,* 2967-2974.

Green, K. P., Kuhl, P. K., Meltzoff, A. N., & Stevens, E. B. (1991). Integrating Speech Information Across Talkers, Gender, and Sensory Modality - Female Faces and Male Voices in the McGurk Effect. *Perception & Psychophysics, 50,* 524-536.

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences, 10,* 14-23.

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

Hamsberger, J. D., Shrivastav, R., Brown, W. S., Rothman, H., & Hollien, H. (2008). Speaking rate and fundamental frequency as speech cues to perceived age. *Journal of Voice, 22,* 58-69.

Hanley, J. R., Smith, S. T., & Hadfield, J. (1998). I recognise you but I can't place you: An investigation of familiar-only experiences during tests of voice and face recognition. *Quarterly Journal of Experimental Psychology Section A-Human Experimental Psychology, 51,* 179-195.

Hanley, J. R. & Turner, J. M. (2000). Why are familiar-only experiences more frequent for voices than for faces? *Quarterly Journal of Experimental Psychology Section A-Human Experimental Psychology, 53,* 1105-1116.

Hari, R., Aittoniemi, K., Jarvinen, M. L., Katila, T., & Varpula, T. (1980). Auditory Evoked Transient and Sustained Magnetic-Fields of the Human-Brain - Localization of Neural Generators. *Experimental Brain Research, 40,* 237-240.

Hartman, D. E. & Danhauer, J. L. (1975). Perceptual Features of Aging Male Speech. *Journal of the Acoustical Society of America, 58,* S92-S93.

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

Holt, L. L. (2006). Speech categorization in context: Joint effects of nonspeech and speech precursors. *Journal of the Acoustical Society of America, 119,* 4016-4026.

Hughes, S. M., Dispenza, F., & Gallup, G. G. (2004). Ratings of voice attractiveness predict sexual behavior and body configuration. *Evolution and Human Behavior, 25,* 295-304.

Imaizumi, S., Mori, K., Kiritani, S., Kawashima, R., Sugiura, M., Fukuda, H. et al. (1997). Vocal identification of speaker and emotion activates different brain regions. *Neuroreport, 8,* 2809-2812.

Jääskeläinen, I. P., Ahveninen, J., Belliveau, J. W., Raij, T., & Sams, M. (2007). Short-term plasticity in auditory cognition. *Trends in Neurosciences, 30,* 653-661.

Jääskeläinen, I. P., Ojanen, V., Ahveninen, J., Auranen, T., Levanen, S., Mottonen, R. et al. (2004). Adaptation of neuromagnetic N1 responses to phonetic stimuli by visual speech in humans. *Neuroreport, 15,* 2741-2744.

Jaquet, E. & Rhodes, G. (2008). Face aftereffects indicate dissociable, but not distinct, coding of male and female faces. *Journal of Experimental Psychology-Human Perception and Performance, 34,* 101-112.

Jaquet, E., Rhodes, G., & Hayward, W. G. (2007). Opposite aftereffects for Chinese and Caucasian faces are selective for social category information and not just physical face differences. *Quarterly Journal of Experimental Psychology, 60,* 1457-1467.

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.

Johnson, R. (1984). P300 - A Model of the Variables Controlling Its Amplitude. *Annals of the New York Academy of Sciences, 425,* 223-229.

Johnston, R. A., Kanazawa, M., Kato, T., & Oda, M. (1997). Exploring the structure of multidimensional face-space: The effects of age and gender. *Visual Cognition, 4,* 39-57.

Johnstone, R. A. (1995). Sexual Selection, Honest Advertisement and the Handicap Principle - Reviewing the Evidence. *Biological Reviews of the Cambridge Philosophical Society, 70,* 1-65.

Jusczyk, P. W. & Luce, P. A. (2002). Speech perception and spoken word recognition: Past and present. *Ear and Hearing, 23,* 2-40.

Kaganovich, N., Francis, A. L., & Melara, R. D. (2006). Electrophysiological evidence for early interaction between talker and linguistic information during speech perception. *Brain Research, 1114,* 161-172.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17,* 4302-4311.

Kausler, D. H. & Puckett, J. M. (1981). Adult Age-Differences in Memory for Sex of Voice. *Journals of Gerontology, 36,* 44-50.

Kawahara, H., Masuda-Katsuse, I., & de Cheveigne, A. (1999). Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds. *Speech Communication, 27,* 187-207.

Kawahara, H. & Matsui, H. (2003). Auditory morphing based on an elastic perceptual distance metric in an interference-free time-frequency representation. *Proceedings of*

the 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol I (Piscataway, NJ: IEEE), 256-259.

Kay, R. H. & Matthews, D. R. (1972). Existence in Human Auditory Pathways of Channels Selectively Tuned to Modulation Present in Frequency-Modulated Tones. *Journal of Physiology-London, 225,* 657-&.

Kitagawa, N. & Ichihara, S. (2002). Hearing visual motion in depth. *Nature, 416,* 172-174.

Klatt, D. H. & Klatt, L. C. (1990). Analysis, Synthesis, and Perception of Voice Quality Variations Among Female and Male Talkers. *Journal of the Acoustical Society of America, 87,* 820-857.

Kloth, N., Wiese, H., & Schweinberger, S. R. (2010). Effekte des Alters auf die Bestimmung des Geschlechts von Gesichtern. In *52.Tagung experimentell arbeitender Psychologen (TeaP), March 2010, Saarbrücken.*

Knösche, T. R., Lattner, S., Maess, B., Schauer, M., & Friederici, A. D. (2002). Early parallel processing of auditory word and voice information. *Neuroimage, 17,* 1493-1503.

Kovacs, G., Zimmer, M., Banko, E., Harza, I., Antal, A., & Vidnyanszky, Z. (2006). Electrophysiological correlates of visual adaptation to faces and body parts in humans. *Cerebral Cortex, 16,* 742-753.

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

Kreiman, J. (1997). Theory and Practice in Voice Perception Research. In K.Johnson & J. W. Mullennix (Eds.), *Talker Variability in Speech Processing* (1 ed., pp. 85-108). San Diego, CA: Academic Press, Inc.

Kreiman, J., Gerratt, B. R., & Precoda, K. (1990). Listener Experience and Perception of Voice Quality. *Journal of Speech and Hearing Research, 33,* 103-115.

Kuriki, S., Ohta, K., & Koyama, S. (2007). Persistent responsiveness of long-latency auditory cortical activities in response to repeated stimuli of musical timbre and vowel sounds. *Cerebral Cortex, 17,* 2725-2732.

Lass, N. J., Hughes, K. R., Bowyer, M. D., Waters, L. T., & Bourne, V. T. (1976). Speaker Sex Identification from Voiced, Whispered, and Filtered Isolated Vowels. *Journal of the Acoustical Society of America, 59,* 675-678.

Latinus, M. & Belin, P. (2010). Adaptation to voice caricatures and anti-voices revealed prototype-based coding of voice identity. In *Cognitive Neuroscience Society - 2010 Annual Meeting* Montreal, Canada.

Lattner, S., Maess, B., Wang, Y. H., Schauer, M., Alter, K., & Friederici, A. D. (2003). Dissociation of human and computer voices in the brain: Evidence for a preattentive gestalt-like perception. *Human Brain Mapping, 20,* 13-21.

Lattner, S., Meyer, M. E., & Friederici, A. D. (2005). Voice perception: Sex, pitch, and the right hemisphere. *Human Brain Mapping, 24,* 11-20.

Lavner, Y., Gath, I., & Rosenhouse, J. (2000). The effects of acoustic modifications on the identification of familiar voices speaking isolated vowels. *Speech Communication, 30,* 9-26.

Lavner, Y., Rosenhouse, J., & Gath, I. (2001). The Prototype Model in Speaker Identification by Human Listeners. *International Journal of Speech and Technology, 4,* 63-74.

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

Levy, D. A., Granot, R., & Bentin, S. (2001). Processing specificity for human voice stimuli: electrophysiological evidence. *Neuroreport, 12,* 2653-2657.

Levy, D. A., Granot, R., & Bentin, S. (2003). Neural sensitivity to human voices: ERP evidence of task and attentional influences. *Psychophysiology, 40,* 291-305.

Lewis, J. W., Wightman, F. L., Brefczynski, J. A., Phinney, R. E., Binder, J. R., & Deyoe, E. A. (2004). Human brain regions involved in recognizing environmental sounds. *Cerebral Cortex, 14,* 1008-1021.

Linville, S. E. (1996). The sound of senescence. *Journal of Voice, 10,* 190-200.

Linville, S. E. & Fisher, H. B. (1985). Acoustic Characteristics of Perceived Versus Actual Vocal Age in Controlled Phonation by Adult Females. *Journal of the Acoustical Society of America, 78,* 40-48.

Linville, S. E. & Korabic, E. W. (1986). Elderly Listeners Estimates of Vocal Age in Adult Females. *Journal of the Acoustical Society of America, 80,* 692-694.

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

Loh, M., Schmid, G., Deco, G., & Ziegler, W. (2010). Audiovisual Matching in Speech and Nonspeech Sounds: A Neurodynamical Model. *Journal of Cognitive Neuroscience, 22,* 240-247.

Mark, V. W. & Stegman, B. A. (1991). Phonagnosia Following Viral Encephalitis. *Journal of Clinical and Experimental Neuropsychology, 13,* 87.

Mathevon, N., Charrier, I., & Aubin, T. (2004). A memory like a female Fur Seal: long-lasting recognition of pup's voice by mothers. *Anais da Academia Brasileira de Ciencias, 76,* 237-241.

May, P., Tiitinen, H., Ilmoniemi, R. J., Nyman, G., Taylor, J. G., & Naatanen, R. (1999). Frequency change detection in human auditory cortex. *Journal of Computational Neuroscience, 6,* 99-120.

McGurk, H. & MacDonald, J. (1976). Hearing Lips and Seeing Voices. *Nature, 264,* 746-748.

Morse, P. A., Kass, J. E., & Turkienicz, R. (1976). Selective Adaptation of Vowels. *Perception & Psychophysics, 19,* 137-143.

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

Mullennix, J. W. & Pisoni, D. B. (1990). Stimulus Variability and Processing Dependencies in Speech-Perception. *Perception & Psychophysics, 47,* 379-390.

Näätänen, R. & Picton, T. (1987). The N1 Wave of the Human Electric and Magnetic Response to Sound - A Review and An Analysis of the Component Structure. *Psychophysiology, 24,* 375-425.

Näätänen, R., Sams, M., Alho, K., Paavilainen, P., Reinikainen, K., & Sokolov, E. N. (1988). Frequency and Location Specificity of the Human Vertex N1-Wave. *Electroencephalography and Clinical Neurophysiology, 69,* 523-531.

Nakamura, K., Kawashima, R., Sugiura, M., Kato, T., Nakamura, A., Hatano, K. et al. (2001). Neural substrates for recognition of familiar voices: a PET study. *Neuropsychologia, 39,* 1047-1054.

Neuner, F. & Schweinberger, S. R. (2000). Neuropsychological impairments in the recognition of faces, voices, and personal names. *Brain and Cognition, 44,* 342-366.

Newman, R. S. & Evers, S. (2007). The effect of talker familiarity on stream segregation. *Journal of Phonetics, 35,* 85-103.

Nygaard, L. C. & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics, 60,* 355-376.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech-Perception As A Talker-Contingent Process. *Psychological Science, 5,* 42-46.

Ockleford, E. M., Vince, M. A., Layton, C., & Reader, M. R. (1988). Responses of Neonates to Parents and Others Voices. *Early Human Development, 18,* 27-36.

Olsson, N., Juslin, P., & Winman, A. (1998). Realism of confidence in earwitness versus eyewitness identification. *Journal of Experimental Psychology-Applied, 4,* 101-118.

Owren, M. J., Berkowitz, M., & Bachorowski, J. A. (2007). Listeners judge talker sex more efficiently from male than from female vowels. *Perception & Psychophysics, 69,* 930-941.

Papcun, G., Kreiman, J., & Davis, A. (1989). Long-Term-Memory for Unfamiliar Voices. *Journal of the Acoustical Society of America, 85,* 913-925.

Perrachione, T. K., Chiao, J. Y., & Wong, P. C. M. (2010). Asymmetric cultural effects on perceptual expertise underlie an own-race bias for voices. *Cognition, 114,* 42-55.

Perrachione, T. K., Pierrehumbert, J. B., & Wong, P. C. M. (2009). Differential Neural Contributions to Native- and Foreign-Language Talker Identification. *Journal of Experimental Psychology-Human Perception and Performance, 35,* 1950-1960.

Perrachione, T. K. & Wong, P. C. M. (2007). Learning to recognize speakers of a non-native language: Implications for the functional organization of human auditory cortex. *Neuropsychologia, 45,* 1899-1910.

Petkov, C. I., Logothetis, N. K., & Obleser, J. (2009). Where Are the Human Speech and Voice Regions, and Do Other Animals Have Anything Like Them? *Neuroscientist, 15,* 419-429.

Pisoni, D. B. (1997). "Normalization" in Speech Perception. In K.Johnson & J. W. Mullennix (Eds.), *Talker Variability in Speech Processing* (1 ed., pp. 9-32). San Diego, CA: Academic Press, Inc.

Pollack, I., Pickett, J. M., & Sumby, W. H. (1954). On the Identification of Speakers by Voice. *Journal of the Acoustical Society of America, 26,* 403-406.

Ragot, R. & LepaulErcole, R. (1996). Brain potentials as objective indexes of auditory pitch extraction from harmonics. *Neuroreport, 7,* 905-909.

Remez, R. E., Fellowes, J. M., & Rubin, P. E. (1997). Talker identification based on phonetic information. *Journal of Experimental Psychology-Human Perception and Performance, 23,* 651-666.

Rendall, D., Rodman, P. S., & Emond, R. E. (1996). Vocal recognition of individuals and kin in free-ranging rhesus monkeys. *Animal Behaviour, 51,* 1007-1015.

Rhodes, G. & Jeffery, L. (2006). Adaptive norm-based coding of facial identity. *Vision Research, 46,* 2977-2987.

Rhodes, G., Jeffery, L., Watson, T. L., Jaquet, E., Winkler, C., & Clifford, C. W. G. (2004). Orientation-contingent face aftereffects and implications for face-coding mechanisms. *Current Biology, 14,* 2119-2123.

Robertson, D. M. C. & Schweinberger, S. R. (2010). The role of audiovisual asynchrony in person recognition. *Quarterly Journal of Experimental Psychology, 63,* 23-30.

Robson, M. D., Dorosz, J. L., & Gore, J. C. (1998). Measurements of the temporal fMRI response of the human auditory cortex to trains of tones. *Neuroimage, 7,* 185-198.

Ryan, W. J. & Burk, K. W. (1974). Perceptual and Acoustic Correlates of Aging in Speech of Males. *Journal of Communication Disorders, 7,* 181-192.

Ryu, J. J., Borrmann, K., & Chaudhuri, A. (2008). Imagine Jane and Identify John: Face Identity Aftereffects Induced by Imagined Faces. *Plos One, 3*.

Saxton, T. K., DeBruine, L. M., Jones, B. C., Little, A. C., & Roberts, S. C. (2009). Face and voice attractiveness judgments change during adolescence. *Evolution and Human Behavior, 30,* 398-408.

Scherg, M. & Voncramon, D. (1985). 2 Bilateral Sources of the Late Aep As Identified by A Spatio-Temporal Dipole Model. *Electroencephalography and Clinical Neurophysiology, 62,* 32-44.

Schirmer, A., Striano, T., & Friederici, A. D. (2005). Sex differences in the preattentive processing of vocal emotional expressions. *Neuroreport, 16,* 635-639.

Schmid, G. & Ziegler, W. (2006). Audio-visual matching of speech and non-speech oral gestures in patients with aphasia and apraxia of speech. *Neuropsychologia, 44,* 546-555.

Schmidtnielsen, A. & Stern, K. R. (1985). Identification of Known Voices As A Function of Familiarity and Narrow-Band Coding. *Journal of the Acoustical Society of America, 77,* 658-663.

Schröger, E. (2007). Mismatch negativity - A microphone into auditory memory. *Journal of Psychophysiology, 21,* 138-146.

Schweinberger, S. R. (2001). Human brain potential correlates of voice priming and voice recognition. *Neuropsychologia, 39,* 921-936.

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

Schweinberger, S. R., Herholz, A., & Sommer, W. (1997). Recognizing famous voices: Influence of stimulus duration and different types of retrieval cues. *Journal of Speech Language and Hearing Research, 40,* 453-463.

Schweinberger, S. R., Herholz, A., & Stief, V. (1997). Auditory long-term memory: Repetition priming of voice recognition. *Quarterly Journal of Experimental Psychology Section A-Human Experimental Psychology, 50,* 498-517.

Schweinberger, S. R., Kloth, N., & Jenkins, R. (2007). Are you looking at me? Neural correlates of gaze adaptation. *Neuroreport, 18,* 693-696.

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

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

Schweinberger, S. R., Zäske, R., Walther, C., Golle, J., Kovács, G., & Wiese, H. (2009). Young without Plastic Surgery: Perceptual adaptation to facial age. In *Joint meeting of the Experimental Psychology Society and the Canadian Society for Brain, Behaviour and Cognitive Science (CSBBCS)* York.

Shah, N. J., Marshall, J. C., Zafiris, O., Schwab, A., Zilles, K., Markowitsch, H. J. et al. (2001). The neural correlates of person familiarity - A functional magnetic resonance imaging study with clinical implications. *Brain, 124,* 804-815.

Sheffert, S. M. & Olson, E. (2004). Audiovisual speech facilitates voice learning. *Perception & Psychophysics, 66,* 352-362.

Sheffert, S. M., Pisoni, D. B., Fellowes, J. M., & Remez, R. E. (2002). Learning to recognize talkers from natural, sinewave, and reversed speech samples. *Journal of Experimental Psychology-Human Perception and Performance, 28,* 1447-1469.

Shipp, T. & Hollien, H. (1969). Perception of Aging Male Voice. *Journal of Speech and Hearing Research, 12,* 703-&.

Shipp, T. & Hollien, H. (1972). Speech Frequency and Duration Measures As A Function of Chronologic Age. *Journal of the Acoustical Society of America, 51,* 111-&.

Shipp, T., Qi, Y. Y., Huntley, R., & Hollien, H. (1992). Acoustic and Temporal Correlates of Perceived Age. *Journal of Voice, 6,* 211-216.

Smith, D. R. R. & Patterson, R. D. (2005). The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age. *Journal of the Acoustical Society of America, 118,* 3177-3186.

Smith, D. R. R., Walters, T. C., & Patterson, R. D. (2007). Discrimination of speaker sex and size when glottal-pulse rate and vocal-tract length are controlled. *Journal of the Acoustical Society of America, 122,* 3628-3639.

Sokhi, D. S., Hunter, M. D., Wilkinson, I. D., & Woodruff, P. W. R. (2005). Male and female voices activate distinct regions in the male brain. *Neuroimage, 27,* 572-578.

Spreckelmeyer, K. N., Kutas, M., Urbach, T., Altenmuller, E., & Muller, T. F. (2009). Neural processing of vocal emotion and identity. *Brain and Cognition, 69,* 121-126.

Strube, G., Goggin, J. P., & Thompson, C. P. (1988). The Role of Comprehension in Voice Identification. *Bulletin of the Psychonomic Society, 26,* 492.

Stufflebeam, S. M., Poeppel, D., Rowley, H. A., & Roberts, T. P. L. (1998). Peri-threshold encoding of stimulus frequency and intensity in the M100 latency. *Neuroreport, 9,* 91-94.

Tesink, C. M. J. Y., Petersson, K. M., van Berkum, J. J. A., van den Brink, D., Buitelaar, J. K., & Hagoort, P. (2009). Unification of Speaker and Meaning in Language Comprehension: An fMRI Study. *Journal of Cognitive Neuroscience, 21,* 2085-2099.

Thompson, C. P. (1987). A Language Effect in Voice Identification. *Applied Cognitive Psychology, 1,* 121-131.

Titova, N. & Näätänen, R. (2001). Preattentive voice discrimination by the human brain as indexed by the mismatch negativity. *Neuroscience Letters, 308,* 63-65.

Torre, P. & Barlow, J. A. (2009). Age-related changes in acoustic characteristics of adult speech. *Journal of Communication Disorders, 42,* 324-333.

Torriani, M. V. G., Vannoni, E., & McElligott, A. G. (2006). Mother-young recognition in an ungulate hider species: A unidirectional process. *American Naturalist, 168,* 412-420.

Tsao, D. Y. & Freiwald, W. A. (2006). What's so special about the average face? *Trends in Cognitive Sciences, 10,* 391-393.

Valentine, T. (1991). A Unified Account of the Effects of Distinctiveness, Inversion, and Race in Face Recognition. *Quarterly Journal of Experimental Psychology Section A-Human Experimental Psychology, 43,* 161-204.

Vanlancker, D. & Kreiman, J. (1987). Voice Discrimination and Recognition Are Separate Abilities. *Neuropsychologia, 25,* 829-834.

Vanlancker, D., Kreiman, J., & Emmorey, K. (1985a). Familiar Voice Recognition - Patterns and Parameters .1. Recognition of Backward Voices. *Journal of Phonetics, 13,* 19-38.

Vanlancker, D., Kreiman, J., & Wickens, T. D. (1985b). Familiar Voice Recognition - Patterns and Parameters .2. Recognition of Rate-Altered Voices. *Journal of Phonetics, 13,* 39-52.

Vanlancker, D. R., Cummings, J. L., Kreiman, J., & Dobkin, B. H. (1988). Phonagnosia - A Dissociation Between Familiar and Unfamiliar Voices. *Cortex, 24,* 195-209.

Vanlancker, D. R., Kreiman, J., & Cummings, J. (1989). Voice Perception Deficits - Neuroanatomical Correlates of Phonagnosia. *Journal of Clinical and Experimental Neuropsychology, 11,* 665-674.

Vaughan, H. G. & Ritter, W. (1970). Sources of Auditory Evoked Responses Recorded from Human Scalp. *Electroencephalography and Clinical Neurophysiology, 28,* 360-&.

von Kriegstein, K. & Giraud, A. L. (2004). Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *Neuroimage, 22,* 948-955.

von Kriegstein, K. & Giraud, A. L. (2006). Implicit multisensory associations influence voice recognition. *Plos Biology, 4,* 1809-1820.

von Kriegstein, K., Kleinschmidt, A., Sterzer, P., & Giraud, A. L. (2005). Interaction of face and voice areas during speaker recognition. *Journal of Cognitive Neuroscience, 17,* 367-376.

Walker, S., Bruce, V., & Omalley, C. (1995). Facial Identity and Facial Speech Processing - Familiar Faces and Voices in the Mcgurk Effect. *Perception & Psychophysics, 57,* 1124-1133.

Warren, J. D., Scott, S. K., Price, C. J., & Griffiths, T. D. (2006). Human brain mechanisms for the early analysis of voices. *Neuroimage, 31,* 1389-1397.

Watson, R. (2009). Selectivity for Conspecific Vocalizations within the Primate Insular Cortex. *Journal of Neuroscience, 29,* 6769-6770.

Webster, M. A., Kaping, D., Mizokami, Y., & Duhamel, P. (2004). Adaptation to natural facial categories. *Nature, 428,* 557-561.

Webster, M. A. & Maclin, O. H. (1999). Figural aftereffects in the perception of faces. *Psychonomic Bulletin & Review, 6,* 647-653.

Wood, C. C. & Wolpaw, J. R. (1982). Scalp Distribution of Human Auditory Evoked-Potentials .2. Evidence for Overlapping Sources and Involvement of Auditory-Cortex. *Electroencephalography and Clinical Neurophysiology, 54,* 25-38.

Woods, D. L., Alho, K., & Algazi, A. (1993). Intermodal Selective Attention - Evidence for Processing in Tonotopic Auditory Fields. *Psychophysiology, 30,* 287-295.

Yonan, C. A. & Sommers, M. S. (2000). The effects of talker familiarity on spoken word identification in younger and older listeners. *Psychology and Aging, 15,* 88-99.

Zahavi, A. (1977). Cost of Honesty - (Further Remarks on Handicap Principle). *Journal of Theoretical Biology, 67,* 603-605.

Zäske, R. & Schweinberger, S. R. (2010). Gender-contingent aftereffects of vocal age adaptation. In *Research Seminars in Psychology and Cognitive Neuroscience, June 2010, Jena, Germany.*

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.

Zäske, R., Schweinberger, S. R., & Kawahara, H. Voice aftereffects of adaptation to speaker identity. *Hearing Research,* (in press).

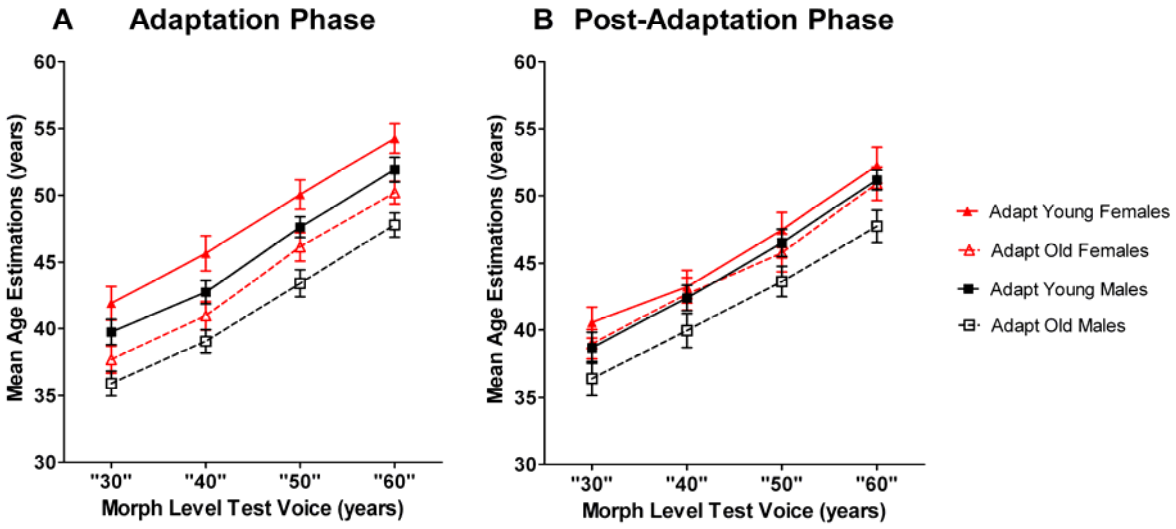
List of Abbreviations

AEP	auditory evoked potential
EEG	electroencephalogram
ERP	event-related potential
Exp.	experiment
F0	fundamental frequency
F1,F2,…Fn	formant frequencies
FAAE	face age aftereffect
FFA	fusiform face area
FIAE	face identity aftereffect
fMRI	functional magnetic resonance imaging
FRU	face recognition unit
Hz	Hertz
m	magnetic
MEG	magnetoencephalogram
MMN	mismatch negativity
ms	milliseconds
MTG	middle temporal gyrus
N	negativity
P	positivity
PET	positron emission tomography
PIN	person identity node
PSE	point of subjective equality
s	seconds
SEM	standard error of the mean
STG	superior temporal gyrus
STRAIGHT	Speech Transformation and Representation using Adaptive Interpolation of weiGHTed spectrum
STS	superior temporal sulcus
TVA	temporal voice area
VAAE	voice age aftereffect
VCV	vowel consonant vowel
VGAE	voice gender aftereffect

VIAE	voice identity aftereffect
VPP	vertex positive potential
VRU	voice recognition unit
yrs	years

Appendix 1

Male Voices



Female Voices

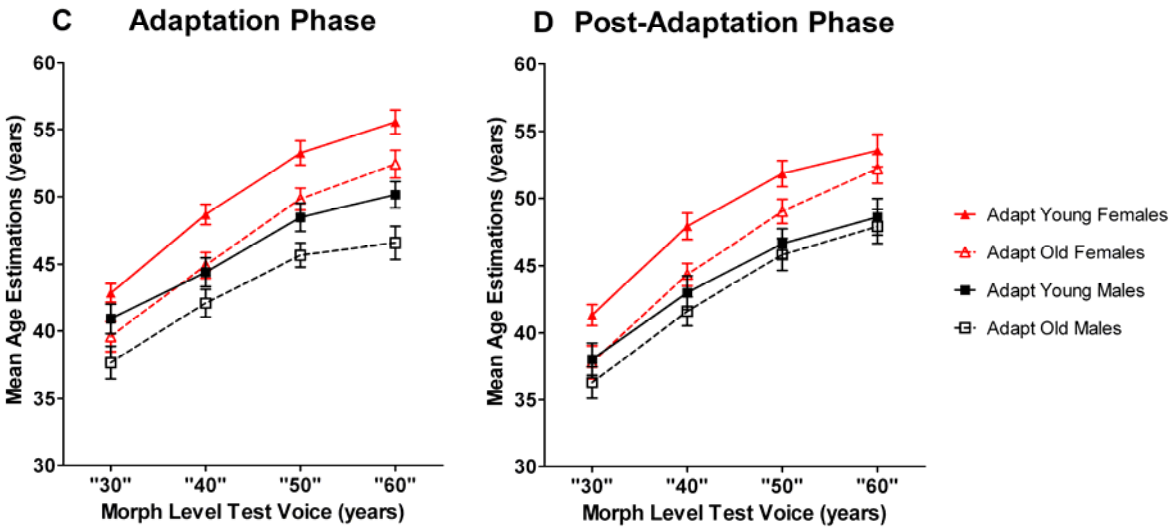


Fig. 3. Vocal age aftereffects (VAAEs) in adaptation phases (A and C) and post-adaptation phases (B and D) induced by old or young voices as a function of speaker and listener gender. Red lines represent female listeners and black lines represent male listeners. Therefore, graphs showing conditions of incongruent (i.e. different) speaker and listener genders are red for male voices (A and B) and black for female voices (C and D). The reverse colour scheme applies to conditions of congruent (i.e. same) speaker and listener gender. Note that in post-adaptation blocks, the VAAE was still measurable, but significantly reduced relative to adaptation blocks when speakers and listeners had different genders. Error bars represent standard errors of the mean (SEM).

Appendix 2

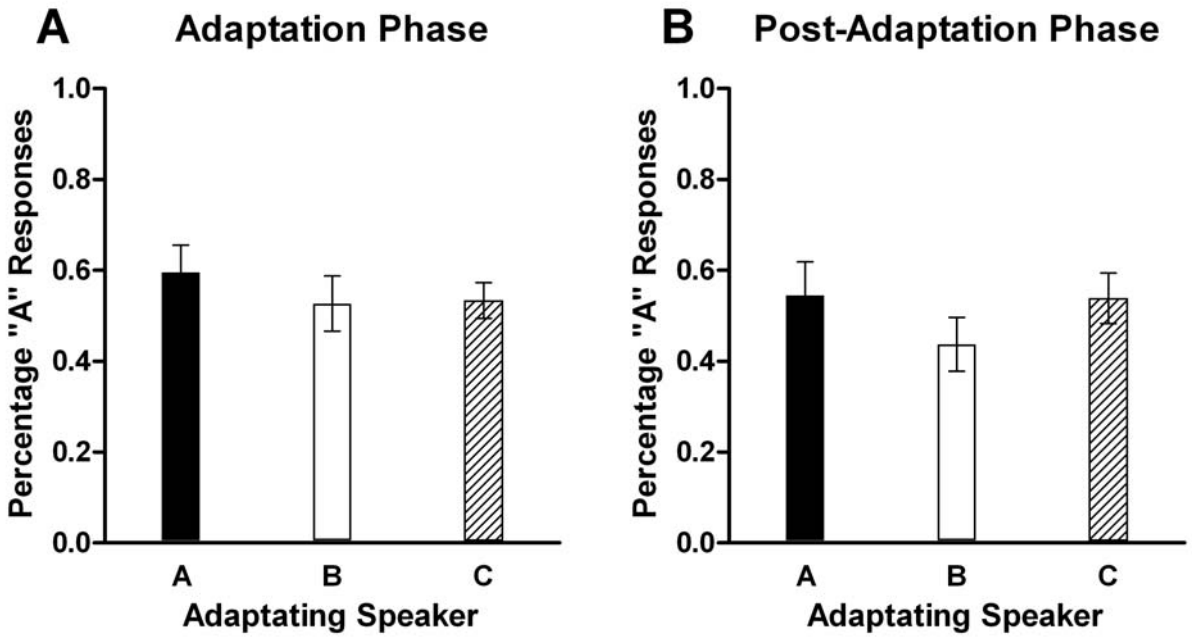


Fig. 2. Percentage of "speaker A" responses to test voices following adaptation to names of speakers A, B, and C in adaptation blocks (A) and in post-adaptation blocks (B). Error bars show standard errors of the mean (SEM).

Ehrenwörtliche Erklärung

Ich erkläre, dass mir die geltende Promotionsordnung der Fakultät für Sozial- und Verhaltenswissenschaften der Friedrich-Schiller-Universität Jena bekannt ist.

Ferner erkläre ich, dass ich die vorliegende Dissertation selbstständig ohne die Hilfe Dritter angefertigt habe, sowie alle benutzten Quellen und Hilfsmittel in der Arbeit angegeben habe. Insbesondere habe ich keine Hilfe eines Promotionsberaters in Anspruch genommen.

Bei der Auswahl und Auswertung des Materials sowie der Herstellung der Einzelmanuskripte haben mich die angegebenen Koautoren unentgeltlich unterstützt. Darüber hinaus hat kein Dritter unmittelbar oder mittelbar geldwerte Leistungen von mir für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Ich erkläre weiterhin, dass ich diese Dissertation noch nicht als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung, oder eine gleiche, eine in wesentlichen Teilen ähnliche oder eine andere Abhandlung bei einer anderen Hochschule bzw. anderen Fakultät als Dissertation eingereicht habe. Ich versichere, nach bestem Wissen die reine Wahrheit gesagt und nichts verschwiegen zu haben.

Jena, 15. Juni 2010 _____