Attention, Perception, & Psychophysics (2019) 81:1675–1697
https://doi.org/10.3758/s13414-018-01657-5

Talker-familiarity benefit in non-native recognition memory and word identification: The role of listening conditions and proficiency

Polina Drozdova[1,2] · Roeland van Hout[1] · Odette Scharenborg[1,3]

Published online: 4 March 2019

Abstract

Native listeners benefit from talker familiarity in recognition memory and word identification, especially in adverse listening conditions. The present study addresses the talker familiarity benefit in non-native listening, and the role of listening conditions and listeners' lexical proficiency in the emergence of this benefit. Dutch non-native listeners of English were trained to identify four English talkers over 4 days. Talker familiarity benefit in recognition memory was investigated using a recognition memory task with "old" and "new" words produced by *familiar* and *unfamiliar* talkers presented either in the clear or in noise. Talker familiarity benefit in word identification was investigated by comparing non-native listeners' performances on the first and the last day in identifying words in different noise levels, produced by either a *trained* (included in the voice recognition training) or by an *untrained* talker (not included in the voice recognition training). Non-native listeners demonstrated a talker familiarity benefit in recognition memory, which was modulated by listening conditions and proficiency in the non-native language. No talker familiarity benefit was found in word identification. These results suggest that, similar to native listening, both linguistic and indexical (talker-specific) information influence non-native speech perception. However, this is dependent on the task and type of speech recognition process involved.

Keywords Familiar talker benefit · Non-native speech comprehension · Noise · Non-native proficiency

Introduction

A speech signal is extremely variable, and not only contains linguistic but also so-called indexical information (Abercombie, 1967). This includes, for instance, information about a talker's age, gender, emotional state, dialect, and accent. Previous studies demonstrated that listeners are sensitive to changes in indexical information: words are better recognized when they are spoken by one talker than when they are spoken by multiple talkers (Mullennix et al., 1989; Ryalls & Pisoni, 1997; Sommers et al., 1994). Moreover, words repeated by the same talker and even words combined from phonemes repeated by the same talker are recognized more quickly and more accurately than words repeated by a different talker (Bradlow & Pisoni, 1999; Cooper & Bradlow, 2017; Goh, 2005; Goldinger, 1996; Jesse et al., 2007; Luce & Lyons, 1998; Palmeri et al., 1993; Sheffert, 1998), indicating that surface details of words, such as indexical information, are retained in some form in the memory of the listeners and subsequently facilitate speech processing.

Additional support for the role of indexical information in speech perception comes from studies demonstrating that talker familiarity facilitates perception of words and sentences. Listeners were shown to be able to accommodate a talker's ambiguous or accented pronunciation during exposure and process subsequent stimuli in a talker-specific manner (Clarke & Garrett, 2004; Dahan et al., 2008; Norris et al., 2003; Trude & Brown-Schmidt, 2012). Nygaard et al. (2008) furthermore showed that listeners are faster at repeating

✉ Odette Scharenborg
O.E.Scharenborg@tudelft.nl

Polina Drozdova
pdrozdova@diakhuis.nl

Roeland van Hout
r.vanhout@let.ru.nl

[1] Centre for Language Studies, Radboud University Nijmegen, Erasmusplein 1, 6525 HT Nijmegen, The Netherlands

[2] Concerndienst, Diakonessenhuis Utrecht, Postbus 80250, 3508TG Utrecht, The Netherlands

[3] Multimedia Computing Group, Delft University of Technology, Van Mourik Broekmanweg 6, 2628 XE Delft, The Netherlands

words when these words were produced by familiar than unfamiliar talkers. The same finding was reported by Maibauer et al. (2014) using famous talkers familiar to the listeners.

The effect of indexical information has been investigated at different levels of the speech recognition process with mixed results. Facilitatory effects of indexical information have been observed in tasks involving recognition memory, addressing sound processing (i.e., in a recognition memory task). In this task type, listeners have to decide whether they have previously heard a word they are hearing now, either in an earlier trial in a continuous recognition memory task (Palmeri et al., 1993) or in an exposure phase in an exposure-test set-up (e.g., Luce & Lyons, 1998). In both cases, listeners were shown to recognize the word faster and more accurately when the word was spoken by the same talker in an earlier presentation of the word. Results obtained with tasks involving lexical access, such as word identification where listeners have to type in the word they heard, and lexical decision, where listeners have to indicate whether a stimulus is an existing word, were inconsistent. Luce and Lyons (1998) found an effect of indexical information in a recognition memory task, but not in a lexical decision task with the same stimuli. However, using an eye-tracking paradigm, Papesh et al. (2016) did find an effect of indexical information in a lexical decision task, while Creel et al. (2008) showed that indexical information is involved in the lexical disambiguation process, i.e., similar sounding words like “couch” and “cow” were found to be activated longer when the listener previously heard them produced by the same talker than when they were produced by different talkers.

Luce et al. (2003) suggested that inconsistencies between studies on indexical effects can be explained by differences in processing time required to perform the task. According to this hypothesis, referred to as the *time-course hypothesis*, the indexical effects are more likely to emerge when processing is slowed down through an increase of task difficulty. The time-course hypothesis was corroborated by findings in tasks focusing on both recognition memory and lexical processing. Mattys and Liss (2008) used normal and dysarthric (mild or severely impaired) speech in a recognition memory task. They found that listeners were faster at recognizing target words when these words were produced by the same talkers than by different talkers, and this difference was larger in the condition with dysarthric speech than in the condition with normal speech. The facilitatory effect of indexical information in a lexical decision task was also shown to depend on task difficulty: indexical effects emerged when word-like non-word stimuli (McLennan & Luce, 2005), foreign-accented speech (McLennan & González, 2012) or low-frequency words (Dufour & Nguyen, 2014; Dufour et al., 2017) were used.

The presence of background noise slows down the speech recognition process (Brouwer & Bradlow, 2011, 2016; Hintz & Scharenborg, 2016). In agreement with the *time-course hypothesis*, indexical information has been found to be accessed in both a recognition memory task and a word identification task when words were embedded in noise (Goldinger, 1996). Further, Nygaard and Pisoni (1998) showed that identification of words in noise is better when these words were produced by familiar talkers (whom the listeners were trained to recognize over the course of 10 days) than when these words were produced by unfamiliar talkers (see also Levi, 2014; Nygaard et al., 1994; Yonan & Sommers, 2000). This talker familiarity benefit was more pronounced for the most difficult noise levels. Finally, Nijveld et al. (2015) found a facilitatory effect of indexical information in a lexical decision task when the words were embedded in noise, but not when the words were in the clear. Note, however, that in that study reaction times for the words in noise were shorter than for the words in the clear, indicating that other factors rather than speed of processing played a role in the emergence of the effects of indexical information. The other factors suggested to play a role are attention to the voice of the talker during encoding (Theodore et al., 2015) or increased attention to the stimuli due to the use of famous voices (Maibauer et al., 2014) or curse words (Tuft et al., 2018).

While native listeners have been repeatedly shown to use indexical information in recognition memory tasks and in tasks involving lexical access when degraded stimuli are used, studies investigating the effect of indexical information on non-native speech processing are scarce. Perceiving speech in a non-native language is more difficult than in a native language due to the mismatch in sound categories between the native and non-native language of the listeners, which leads to spurious activation of candidate words from both the native and non-native language during word recognition (e.g., Broersma, 2012; Weber & Cutler, 2004; Scharenborg et al., 2018). Additionally, listening in the presence of noise is more challenging for non-native than for native listeners (e.g., Mayo et al., 1997; Rogers et al., 2006; Scharenborg et al., 2018; see Garcia Lecumberri et al. (2010) for a review). Despite the non-native listeners’ impaired sound perception, previous research has shown that non-native listeners are sensitive to indexical information in the speech signal at least in clear listening conditions. English learners of German, in a recognition memory task, were found to correctly recognize more German words as already presented when they were repeated by the same talker than when they were repeated by a different talker (Winters et al., 2013).

Similarly, non-native listeners of Spanish, tested in a word repetition task, were only faster at repeating already presented words than new words when these words were produced by the same talker as during the exposure phase (Trofimovich, 2005). This facilitation was shown to be dependent on the listeners' amount of experience with the non-native language (Trofimovich, 2008). Trofimovich (2008) argued that more experienced listeners were more sensitive to phonetic detail in spoken words in the non-native language, and therefore, experienced more facilitation from the same talker than the listeners with less experience in the non-native language. Further, similar to native listeners, non-native listeners were shown to recognize words better when they were produced by one talker than when they were produced by multiple talkers (Bradlow & Pisoni, 1999; Tamati & Pisoni, 2014), and adapted to the accent of previously unfamiliar talkers and used this knowledge in a subsequent recognition of words from these talkers (Drozdova et al., 2016; Reinisch et al., 2013). Finally, a number of studies demonstrated that non-native listeners are able to learn to recognize previously unfamiliar talkers speaking in a non-native language (Drozdova et al., 2017; Bregman & Creel, 2014; Perrachione & Wong, 2007). These studies demonstrate that talker-specific information is stored in the memory of non-native listeners, similar to what is observed in native listeners, and that this information facilitates their recognition memory.

Evidence whether indexical information facilitates non-native word identification only comes, to the best of our knowledge, from one study so far. Levi and colleagues (Levi et al., 2011) trained listeners to recognize talkers in either their native or an unfamiliar language, and compared their word identification performance with words produced by familiar and unfamiliar talkers. Listeners were trained to identify English–German bilingual speakers when these talkers were speaking either in English (the native language of the listeners) or in German (an unfamiliar language to the listeners). When performing a subsequent English word identification task in noise with the same (as learned in the either English or German training phase) and new talkers, only those listeners who were trained in their native English language benefited from familiarity with the talker, while no talker familiarity benefit was observed for the listeners trained on German speech.

The aim of the present study is to investigate the effect of indexical information and, more specifically, the effect of talker familiarity on recognition memory and word identification when listening in a non-native language in both clear and noisy listening conditions. Focusing on non-native listening also allows us to investigate the possible effects of proficiency in the non-native language on the talker familiarity benefit. Specifically, we aim to answer the following research questions:

1. Do non-native listeners benefit from talker familiarity in recognition memory and word identification?
2. What are the effects of the presence of background noise and proficiency in the non-native language on the talker familiarity benefit?

Previous studies demonstrated that non-native listeners are sensitive to indexical information in their recognition memory (Trofimovich, 2005; Winters et al., 2013), but it is unclear whether the presence of noise interacts with this effect. On the one hand, it is possible that non-native listeners will have difficulty picking up indexical information from the signal due to impaired sound perception (Perrachione & Wong, 2007) which will be worsened by the masking effect of background noise on the foreground speech (see for a review Garcia Lecumberri et al., 2010). On the other hand, given that indexical information was shown to affect native recognition memory particularly when listening conditions are difficult, non-native listeners might benefit from talker familiarity especially when listening conditions are noisy. Regarding word identification, native listeners have previously been shown to use indexical information identifying words in noisy listening conditions (Goldinger, 1996; Nygaard & Pisoni, 1998; Nijveld et al., 2015). Levi and colleagues (2011) hypothesized that listeners need to establish acoustic-phonetic links between talker information and what is being said during the training in order to use talker familiarity in word identification. We therefore hypothesize that non-native listeners who have knowledge of the non-native language, and are thus expected to be able to establish this connection, will have the familiar talker benefit in word identification when background noise is present similar to what was observed for native listeners. Since non-native listeners have less stable, detailed, and abstract lexical and phonetic knowledge than native listeners (Garcia Lecumberri et al., 2010), studying these listeners is a unique test of the interaction between abstract linguistic and talker-specific information, and the usage of talker-specific information in lexical processing.

To answer the research questions, non-native Dutch listeners of English were trained to recognize four previously unfamiliar British English talkers over the course of four days. The Dutch–English language pair allows us to investigate the talker familiarity benefit in non-native recognition memory and word identification in noise with little mismatch at the phonological and sound levels between the two languages involved. These similarities and a general high level of English proficiency of the Dutch speakers ensure that the listeners might establish acoustic-phonetic links between indexical and linguistic information in the signal, and that the presence of noise is not completely detrimental for their understanding of the words. They are expected to be able to familiarize themselves with the talkers and understand the words produced by them: two important conditions for the talker familiarity benefit to emerge (Levi et al., 2011; Nygaard & Pisoni, 1998).

On each training day, these Dutch participants performed a recognition memory task, where words were presented in the clear and embedded in noise. The effect of talker

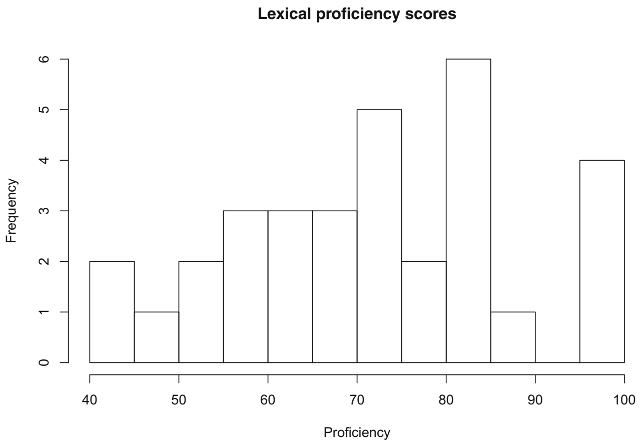

Fig. 1 Breakdown of the lexical proficiency scores of the non-native participants included in this study

familiarity on word identification was studied in a word identification task with various levels of noise. On the first experimental day, listeners' non-native proficiency was assessed.

Method

Participants

Thirty-five native Dutch participants (eight males, M_{age} = 22.4, SD_{age} = 2.34) participated in the experiment. Participants were recruited from the Radboud University Nijmegen subject pool and received either course credits or a monetary reward at the end of the 4-day experiment. This is a subset of the participants described in Drozdova et al. (2017), as not all the participants in that study completed all the tasks necessary for the present study. Prior to the experiment, all participants had to fill in a questionnaire containing questions about their hearing or possible learning disorders. Only those participants indicating no history of hearing or learning disorders were included in the experiment. The average LexTALE score, the test used to measure the listeners' proficiency in English, was 72.1 (SD=17.2) which corresponds to an upper-intermediate level of second language proficiency (Lemhöfer & Broersma, 2012). Figure 1 presents a breakdown of the proficiency scores.

Two additional groups of participants participated in two pre-tests. Fourteen native Dutch participants (two males, M_{age} = 21.71, SD_{age} = 2.02) took part in a word identification task to determine the appropriate noise levels for the experiment, while 16 other participants (two males, M_{age} = 21.94, SD_{age} = 2.84) took part in the pre-test to check the design of the experiment and difficulty levels of the tasks. None of the pre-test participants took part in the main experiment.

Overall design of the experiment

Table 1 shows the overview of the experiment, with the different tasks listed for each of the four days. The different tasks and experiments will be explained in detail in the following subsections. The overall design of the experiment

Table 1 Overview of the experimental tasks per day and the number of words, noise levels, and talkers involved in each task

Day	Tasks	Words	Noise level	Talkers	Duration
1	Word identification	60	−5,0,5 SNR+quiet	A/B	45 min
	Voice recognition training	24+128	quiet	A, C, D, E	
	Recognition memory	32+32	+5 dB SNR+quiet	A, C, D, E + 2 new talkers changed every day	
	LexTale				
2	Voice recognition training	128	clear	A, C, D, E	30 min
	Recognition memory	32+32	+5 dB SNR+clear	A, C, D, E + 2 new talkers	
3	Voice recognition training	128	clear	A, C, D, E	30 min
	Recognition memory	32+32	+5 dB SNR+clear	A, C, D, E + 2 new talkers	
4	Voice recognition training	128	clear	A, C, D, E	45 min
	Recognition memory	32+32	+5 dB SNR+clear	A, C, D, E + 2 new talkers	
	Word identification	60	−5,0,5 SNR+clear	A/B	

The column 'Duration' denotes the total duration of the experimental session for each day was as follows. Day 1 started with a word identification task, in which the participants had to recognize words in the clear and in three noise conditions with different signal-to-noise ratios (SNR) by typing in the word they heard. One group of participants identified words produced by a talker on which they were subsequently trained (*trained* talker condition), while the other group identified the words produced by a talker not included in the voice recognition training (*untrained* talker condition). The word identification task was followed by the voice recognition training where the participants were trained to recognize the voices of four talkers, with a unique combination of talkers for each participant. The voice recognition training was done on all four experimental days and was conducted on words different from the words used in the word identification task. On each experimental day, it was followed by a recognition memory task with the words presented either in the clear or in noise, where listeners had to indicate whether they had previously heard the word in the voice recognition training. The words were either produced by the talkers with which the listeners were familiarized in the voice recognition training (*familiar* talkers), or by *unfamiliar* talkers. The LexTALE task to assess the level of proficiency in the non-native language was carried out at the end of day 1. Finally, at the end of day 4, a second word identification task was conducted.

Talkers

All stimuli were recorded by 13 male native British English speakers, who at the time of the experiment were living (working or studying) in or visiting the Netherlands. Table 2 presents for each talker the average, standard deviation, and range (minimum and maximum) of the F0 as measured by Praat (Boersma & Weenink, 2009) in Hz, and the average word length in ms. Talker 1 was included in all tasks (talker 1 is the *trained* talker in the word identification task and denoted as talker A in Table 1), which was necessary for another experiment not reported here. Talkers 1–12 were included in both the voice recognition training and the recognition memory task. Talker 13 (denoted as talker B in Table 1) was chosen randomly from the list of available talkers, and was only used as the *untrained* talker in the word identification task.

In order to compare the voice characteristics of the 12 different talkers with whom the listeners were familiarized, the average word length and F0 measures were calculated on the basis of the 128 words used in the voice recognition training. To compare the *trained* (talker 1) and *untrained* (talker 13) talkers in the word identification task, the average word length and F0 measures were calculated on the basis of the 60 words from this task. Because of his appearance in both experiments, talker 1 appears in both parts of Table 2.

As can be seen in Table 2, there was a variety of fundamental frequencies (ranging from 90 to 179 Hz) and speaking rates (as indexed by the average word lengths which ranged from 431 to 624 ms). Since a unique combination of talkers was used for each participant in the voice recognition training and the recognition memory task, talker familiarity benefit could be investigated irrespective of how distant or similar the voices were. The role of

Table 2 Characteristics of the talkers used in the experiment

	F0				Word Length (ms)	
Talker	Mean	SD	Minimum	Maximum	Mean	SD
Voice recognition training and recognition memory task						
1	107	15	89	137	585	116
2	98	16	73	129	556	122
3	153	27	115	198	490	106
4	119	13	104	143	527	118
5	90	26	73	142	516	103
6	137	19	116	162	579	137
7	114	20	94	144	437	97
8	148	15	133	174	526	110
9	156	23	103	230	569	141
10	122	21	97	155	624	124
11	115	20	93	145	431	96
12	142	11	126	165	548	119
Word identification task						
1	106	14	90	137	561	113
13	179	13	153	224	564	100

talker-specific characteristics on voice recognition is studied in detail in Drozdova et al. (2017), therefore we will not focus on them in the present study. The talkers were recorded individually in a sound-proof booth with a Sennheiser ME 64 microphone at a sampling frequency of 44100 Hz. Each word was pronounced at least twice by each talker. Words which were mispronounced or produced too quietly were recorded again. The words were then excised from the resulting audio files using a Matlab (The MathWorks Inc., 2013) script, and the segmentations were subsequently manually checked using Praat (Boersma & Weenink, 2009). All talkers were rewarded 5 euros for half an hour of recording time.

Materials, experimental set-up, and procedure

All participants were tested individually in a quiet sound-attenuated booth. The stimuli were presented to them binaurally through headphones. The intensity level of all the stimuli was set at 70 dB SPL. The experiment was administered using Presentation software (Neurobehavioral Systems, Inc., Berkeley, CA, http://www.neurobs.com). In all tasks, participants were asked to react as quickly as possible while trying to avoid making mistakes. All words used in the different parts of the experiment are presented in Appendix.

Voice recognition training

In order to investigate talker familiarity benefit in recognition memory and word identification, it is crucial that participants familiarized themselves with the talkers used in both tasks. To that end, a 4-day voice recognition training was implemented. The training consisted of three phases: a *familiarization*, a *feedback*, and a *test* phase, where the *familiarization* phase was only present on the first day of the experiment.

The voice recognition training included 76 monosyllabic and 76 bisyllabic content words (word frequencies were retrieved from the *SUBTLEX-UK* database (Van Heuven et al., 2014) and ranged from 1.02 per million to 589 per million). Twenty-four words (12 monosyllabic and 12 bisyllabic) were used in the *familiarization* phase on day 1. The remaining 128 words were used for the *feedback* and *test* phases on days 1–4. These words were semi-randomly split over these two phases on each training day (so that each part contained the same number of bisyllabic and monosyllabic words and contained words of comparable frequency). The same words were used on each training day, but their distribution over the *test* and *feedback* phases differed every day. Moreover, the distributions for day 1 through day 4, which were used for half of the participants were reversed for the other half of the participants. Finally, following Nygaard and Pisoni (1998), the talker who produced the word on each day varied (i.e., if a word was produced by talker 1 on day 1 it was, e.g., produced by talker 2 on day 2, etc.). Each participant was trained to recognize four talkers of the set of 12 talkers (talker 1–12), where all participants had to learn the voice of talker 1. To that end, 11 combinations of talkers (lists) were created (e.g., list 1: talker 1, 5, 8, 9; list 2: talker 1, 11, 7, 12, etc.). Each participant was randomly assigned to one of the lists.

In the *familiarization* phase, participants were instructed they would hear words spoken by four different talkers, and their task was to memorize the voice and the name of the talker, which was shown on a computer screen. Participants were presented with five words from each talker, followed by a sequence of four words, each of which was again spoken by one of the four talkers. This procedure was repeated twice. Participants pressed a button on a button box when they were ready to move to the next word. In the *feedback* phase, participants saw the four names of the talkers on the screen. Upon hearing the stimulus, they had to press the button on the button box corresponding to the name of the talker they thought had spoken the word. Subsequently, the participants received feedback in the form of the word "correct" appearing on the screen in case of a correct response or the name of the correct talker in case of an incorrect response. The *test* phase was similar to the feedback phase but without feedback provided to the participants.

Recognition memory task

The design used of the recognition memory task is shown in Table 3. The recognition memory task consisted of four conditions: old talker/old word, old talker/new word, new talker/old word, and new talker/new word. On each training day, the recognition memory task included 64 words, 32 of which were already presented to the participants in the *feedback* or *test* phases during the voice recognition training on that same day (16 from each phase), and 32 were "new" words, the participants had not heard before in the context of the experiment. Note that of the "old" words spoken by the same talker as during the voice recognition training, a different token (i.e., a different rendition) from the one used during the voice recognition training was chosen. The set of words was different for each training day. So, in total, 128 "old" words and 128 "new" words were used in the four recognition memory tasks. The word frequencies in this task ranged from 1.02 per million to 1778 per million (Van Heuven et al., 2014).

The second crucial manipulation was the talker of the "old" and "new" words. Half of the words presented to the participants were spoken by the four talkers on which the listeners were trained. The other 32 words were spoken

Table 3 The number of words per condition in the recognition memory task

	Old talker (4 different talkers)	New talker (2 different talkers)
Old word	16 (*each talker*: 4 words, 2 of which in the clear and 2 in noise)	16 (*each talker*: 8 words, 4 of which in the clear and 4 in noise)
New word	16 (*each talker*: 4 words, 2 of which in the clear and 2 in noise)	16 (*each talker*: 8 words, 4 of which in the clear and 4 in noise)

by two "new", *unfamiliar* talkers, different for each day. The "new" talkers were chosen from the remaining set of 8 talkers (12-4 talkers on which the listener was trained). Finally, half of the words in each condition were presented in speech-shaped noise at an SNR of 5 dB. Following the procedure described in Scharenborg et al. (2018), noise was automatically added to the words using a PRAAT script. Each word was preceded and followed by 200 ms of noise, and 20 ms of lead-in noise was added. Before adding noise the audio file was down-sampled to 16000 Hz to match the sampling frequency of the noise file.

The participants were instructed that they would hear words, some of which they had already heard during the voice recognition training on that day. They were told that some words would be embedded in noise, but were asked not to pay attention to the noise or to the talker who produced the word. The task of the participant was to decide whether the word they heard was already presented to them or whether the word was new. This task, therefore, required explicit recollection of previously heard words. Participants had to indicate their answer by pressing one of two buttons on a button box. To aid the listeners, two options appeared on the screen: "old" corresponded to the left button and appeared on the left side of the screen and "new" corresponded to the right button and appeared on the right side of the screen.

Word identification task

Thirty mono- and 30 bisyllabic words were chosen from the SUBTLEX-UK database (Van Heuven et al., 2014). Word frequencies ranged from 0.2 per million to 977 per million; the average frequencies for the monosyllabic and bisyllabic words were comparable. Four listening conditions were used: one clear listening condition and three conditions with speech-shaped noise at three different SNRs. Each participant was presented with each listening condition and each word occurred only once during the task for each participant. To that end, the set of 60 words was divided into four blocks (so 15 words in each block = listening condition) such that the number of monosyllabic and bisyllabic words and the word frequencies were similar in the four blocks. The listening conditions for each block were randomized across participants. Additionally, two different orders of presentation of the blocks (= two experimental lists) were used. Different renditions for each word by each talker were used on the first and fourth day of the experiment.

Three different noise ratios were used: SNR=5 dB, 0 dB, and -5 dB. Noise was added to the stimuli in the same way as to the stimuli in the recognition memory task. The SNRs were chosen on the basis of a pre-test, in which participants heard words from different talkers (talker 1 and four other randomly chosen talkers from the set of 12 talkers at different noise ratios (-10 dB, -5 dB, 0 dB, 5 dB)) and had to type the words they heard. Since -10 dB appeared to be too difficult for the listeners (overall accuracy below 20% correct), it was decided to use -5 dB as the lowest SNR (overall accuracy of 42%; 50% correct for talker 1).

Familiarity with the talker was a between-subject factor. Participants were randomly assigned to one of the experimental lists and to one of the two talker conditions, i.e., the *trained* (= talker 1) or *untrained* talker (= talker 13) condition. Before the task started, participants were instructed that they would hear words, some of which would be in noise, and they would have to type the word they heard. Each block of 15 words was followed by a pause. To start the next block, participants had to press a key.

Language test (LexTALE)

Proficiency in the non-native language was assessed using a visual unspeeded lexical decision task for advanced learners of English (LexTALE: Lemhöfer and Broersma (2012)). Participants were presented with 60 items (words and non-words) which were shown on a screen one-by-one, and had to indicate by button press whether the item on the screen was an existing word in English.

Results

Because of a technical error, the data from one participant on day 3 for the recognition memory task and from one participant in the voice recognition training were not recorded. Additionally, the LexTALE result of one of the participants was missing. The data of these three participants were excluded from all analyses. Three sets of analyses were carried out. To establish whether participants had improved in the recognition of the four talkers from day 1 to day 4, and specifically in the recognition of talker 1, the first set of analyses investigated the responses during the *test* phase of the voice recognition training.

The second set of analyses compared the responses and reaction times of the listeners in the recognition memory task on the words produced by the *familiar* and *unfamiliar* talkers to investigate the role of talker familiarity on recognition memory. Additionally, the effects of background noise and lexical proficiency were investigated.

The third set of analyses investigated the talker familiarity benefit in word identification by comparing the improvement in word identification performance of the listeners before (on day 1) and after (day 4) the voice recognition training between the *trained* and *untrained* talker conditions. We expected listeners in the *trained* talker condition to show more improvement than listeners in the *untrained* talker condition due to the talker familiarity built up during the four training days.

Voice recognition training

Participants' responses in the *test* phase of the voice recognition training were analyzed using a repeated-measures analysis of variance (ANOVA; following Levi et al., 2011; Nygaard & Pisoni, 1998; Yonan & Sommers, 2000). Each participant was exposed to four talkers, and proportions of hits (correct responses) and false alarms (participant thinks that the word was produced by the target talker while it was produced by another talker) were calculated. Since on each day participants identified 64 stimuli by four talkers, the maximum number of correct responses for each talker was 16, while the maximum number of false alarms was 48. The proportions of hits and false alarms were used to calculate d-prime (d'), a common sensitivity index to measure accuracy performance of participants (Macmillan & Kaplan, 1985) in a recognition task. Voice recognition improvement is then the improvement of sensitivity over time (i.e., the within-subject factor day). Figure 2 illustrates the voice recognition performance in the *test* phase measured with d' for all talkers (gray line) and for talker 1 only (black line), split out per training day.

Due to the word identification task being administered prior to the voice recognition training, listeners in the *trained* talker condition had already been exposed to talker 1 prior to the first voice recognition training session while the listeners in the *untrained* talker condition were not. To account for this difference in exposure and to investigate whether both listener groups were able to learn the voices of the talkers, talker condition (*trained* vs. *untrained*) was included in the analysis as a between-subject factor.

The ANOVA analysis of the voice recognition performance for the voices of all talkers showed a significant difference between experimental days (F (3,93) = 12.01, p <0.001). As can be seen in Fig. 2 (gray line), although voice recognition performance was already high after the first experimental day, listeners demonstrated significant improvement from the first to the last experimental day. Neither the difference in performance between the talker conditions (F (1, 30) = 0.728, p = 0.400) nor the talker condition by day interaction were statistically significant (F (3, 90) = 0.648, p = 0.586). So the greater exposure to the talker in the word identification task for the *trained* talker condition compared to the *untrained* talker condition did not lead to significant differences in voice recognition performance between the two groups.

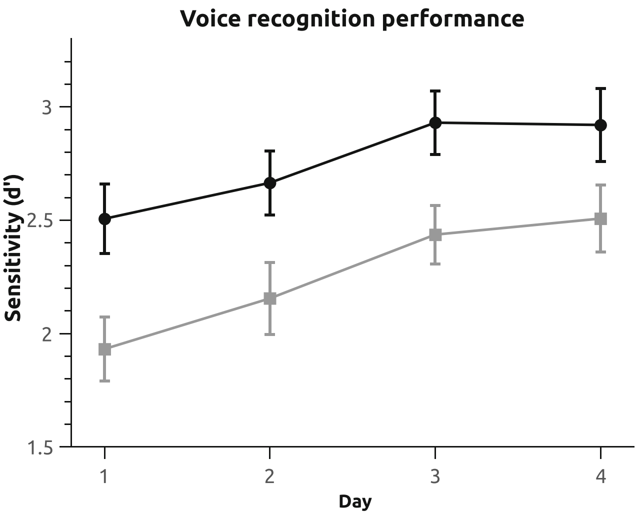

Fig. 2 Voice recognition performance or sensitivity (d') with standard errors for each of the four voice recognition training days. The *gray line* represents voice recognition performance averaged across all words and talkers. The *black line* represents listeners' sensitivity for the voice of talker 1

The improvement in the recognition of talker 1 from day 1 to day 4 has to be taken into account for the evaluation of the performance of the participants in the word identification task. Therefore, listeners' voice recognition performance on the voice of talker 1 was analyzed separately (see the black line in Fig. 2). The statistical analysis showed a significant improvement in the recognition of talker 1 from day 1 to day 4 (F (3, 93) =3.618, p = 0.016). Note that the listeners in the *trained* and *untrained* talker conditions did not differ in their recognition of the voice of talker 1 (talker condition: F (1, 30)=0.099, p=0.814, interaction between talker condition and day: F (3,90) = 2.095, p = 0.107). The additional exposure to talker 1 during the word identification task on day 1 did thus not significantly influence the performance gain during the voice recognition training from day 1 to day 4 in the *trained* talker condition. To summarize, the voice recognition training was successful: participants improved their recognition of the four talkers they were exposed to as well as their recognition of talker 1 irrespective of whether they had been exposed to the voice of talker 1 prior to the voice recognition training.

Talker familiarity in recognition memory

The use of indexical information in recognition memory was measured on two levels. To investigate the role of talker familiarity on recognition memory the sensitivity rates (d') for "old" words were computed for "old" (*familiar*) and "new" (*unfamiliar*) talkers and the reaction times to hits were investigated (cf. Goh, 2005; Luce & Lyons, 1998; Palmeri et al., 1993). Following the studies focusing on the familiar talker benefit in recognition memory (e.g., Goh, 2005; Luce & Lyons, 1998; Palmeri et al., 1993), listeners were expected to be faster and more accurate recognizing words as "old" when they were produced by the same talker as in the exposure (*familiar* talker) than when they were produced by *unfamiliar* talkers.

Additionally, listeners' voice recognition performance (voice recognition accuracy on each testing day for each participant, see "Voice recognition training") was included in the analysis to investigate whether listeners who were more successful in voice recognition benefited from familiarity with the talker to a greater extent than less successful listeners. Nygaard and Pisoni (1998) reported differences between listeners in the size of the talker familiarity benefit for word identification in noise, but no previous studies investigated the role of the degree of familiarity with the talker on the talker familiarity benefit in recognition memory. The inclusion of voice recognition performance as a factor in the analysis allows us to address the question whether individual listeners succeed in using their specific indexical knowledge of the talkers to increase their performance in identifying "old" words from "new" words.

All analyses for the recognition memory task were conducted by means of linear mixed effects models (Jaeger, 2008) using lmer (package lme4) with either accuracy or reaction times as a dependent variable. To investigate the role of listening conditions and lexical proficiency in the emergence of the talker familiarity benefit, noise (present or absent) and lexical proficiency measured with LexTALE were included in all analyses. The analyses were performed in a step-wise manner starting from the most complex model including all the factors of interest and the interactions between them. All continuous factors were scaled and centered. Non-significant factors were removed from the model one by one, starting with non-significant interactions, and comparing each subsequent model with a previous one using the deviance score (-2 * the log-likelihood ratio).

Accuracy

The d' for the "old" words was calculated for each participant per day and listening condition and included in the linear mixed effects model analysis as the dependent variable with noise, talker (*familiar* or *unfamiliar*), voice recognition performance, day (1-4), and lexical proficiency as fixed factors. Subject and list (combination of talkers the participant was exposed to) were added as random factors. Given that participants could differ in the degree of improvement per day, sensitivity for noise, and perception of particular talkers' voices, by-subject random slopes for day, noise and talker were also included as random factors. Day was included as a categorical variable with day 1 as a reference value. Using day as a categorical variable rather than a continuous one enables us to compare the performance of participants on each testing day to their performance on day 1 when the target talker was the least familiar. The p values were obtained by treating the t statistics as z statistics (Barr et al., 2013).

Figure 3 illustrates the d' with standard errors for recognizing "old" words produced by *familiar* (black

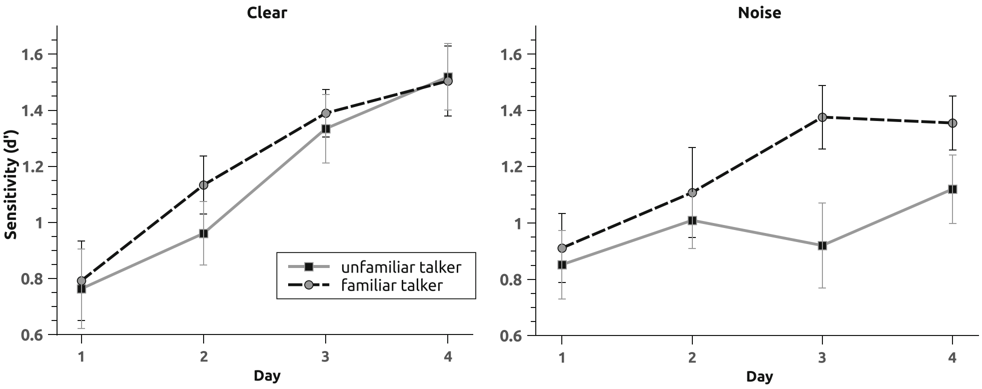

Fig. 3 Sensitivity d' and standard errors of the listeners in the recognition memory task in recognizing "old" words, split out by listening condition and talker (*familiar* vs. *unfamiliar*)

dashed line with bullets) and *unfamiliar* (gray solid line with squares) talkers per day and per listening condition, with the results for the clear listening condition in the left panel and those for the noise condition in the right panel. The estimates from the best fitting model from this analysis are provided in Table 4. This final model only included Subject as a random factor, as other random slopes and intercepts did not significantly improve the model, demonstrating that the accuracy in recognizing "old" words was not influenced by differences in participants' voice recognition improvement per day, their sensitivity to noise, or differences in who the four talkers were in the list.

Importantly, the analysis revealed a general effect of Talker (Table 4: Talker). Its magnitude does not depend on the degree of familiarity with the talker (no significant interaction between voice recognition performance and talker). Additionally, listeners significantly improved their recognition of "old" words on the third and the fourth training days in comparison with the first training day as shown by a significant effect of day. The improvement on the second day was moderated by the voice recognition performance with a larger improvement observed for the listeners with a lower voice recognition performance (interaction between day 2 and voice recognition performance). Listeners with a larger voice recognition performance were in general better at recognizing "old" words, but this effect was weaker for the words in noise (interaction between noise and voice recognition performance). The presence of the factor noise in a number of three-way interactions (i.e., between noise, voice recognition performance, and lexical proficiency, and between noise, voice recognition performance, and day), reveals systematic differences in performance of the listeners on words in the clear and in noise. To investigate these differences between these listening conditions, and to further address the question about the role of noise in the emergence of the familiar talker benefit in recognition memory, two new analyses were carried out, for the clear and noise listening conditions separately. Table 5 provides estimates for the best fitting model for the words in the clear

Table 4 Estimates for the best fitting model for the *d′* measures of the "old" words in the recognition memory task

Fixed effects	*β*	*SE*	*t*	*p*
Intercept	0.928	0.110	8.441	<0.001
Day 2	0.181	0.123	1.476	0.140
Day 3	**0.496**	**0.125**	**3.968**	**<0.001**
Day 4	**0.610**	**0.126**	**4.832**	**<0.001**
Talker	**−0.136**	**0.058**	**−2.232**	**0.020**
Noise	−0.005	0.127	−0.039	0.969
Lexical Proficiency	0.000	0.055	0.000	1.000
Voice recognition performance	**0.184**	**0.080**	**2.289**	**0.022**
Voice recognition performance x noise	**−0.279**	**0.108**	**−2.592**	**0.010**
Day 2 x Noise	0.049	0.173	0.284	0.777
Day 3 x Noise	−0.189	0.175	−1.079	0.281
Day 4 x Noise	−0.193	0.176	−1.094	0.274
Lexical Proficiency x Noise	0.091	0.061	1.499	0.134
Day 2 x voice recognition performance	**−0.215**	**0.102**	**−2.103**	**0.035**
Day 3 x voice recognition performance	−0.170	0.114	−1.494	0.135
Day 4 x voice recognition performance	−0.074	0.106	−0.697	0.486
Lexical proficiency x voice recognition performance	0.021	0.045	0.460	0.645
Noise x lexical proficiency x voice recognition performance	**−0.106**	**0.054**	**−1.943**	**0.052**
Noise x day 2 x voice recognition performance	**0.364**	**0.143**	**2.542**	**0.011**
Noise x day 3 x voice recognition performance	0.286	0.159	1.795	0.073
Noise x day 4 x voice recognition performance	0.141	0.148	0.952	0.341
Random effects			*SD*	
Subject	intercept		0.191	

Significant factors and interactions are presented in bold

Table 5 Estimates for the best fitting model for the d' measures of the “old” words in the clear listening condition in the recognition memory task

Fixed effects	β	*SE*	*t*	*p*
Intercept	0.862	0.093	9.308	<0.001
Day 2	0.180	0.121	1.489	0.137
Day 3	**0.493**	**0.123**	**4.006**	**<0.001**
Day 4	**0.608**	**0.124**	**4.907**	**<0.001**
Voice recognition performance	**0.185**	**0.078**	**2.383**	**0.017**
Day 2 x voice recognition performance	**−0.221**	**0.101**	**−2.185**	**0.029**
Day 3 x voice recognition performance	−0.158	0.118	−1.410	0.158
Day 4 x voice recognition performance	−0.066	0.105	−0.635	0.526
Random effects			*SD*	
Subject	intercept		0.153	

Significant factors and interactions are presented in bold

listening condition, and Table 6 provides estimates for the best fitting model for the words presented in noise.

The analysis for the clear listening condition did not reveal a significant difference in performance for the words produced by *familiar* and *unfamiliar* talkers (the talker effect is not present in Table 5, see also the left panel in Fig. 3). At the same time, listeners who were more accurate in recognizing the voices of the familiar talkers in the voice recognition training were also better in identifying “old” words (factor voice recognition performance), irrespective of whether these words were produced by *familiar* or *unfamiliar* talkers. Given that the words from the voice recognition training are used as “old” words in the recognition memory task, it seems that listeners with a higher voice recognition performance might have been able to establish the connection between the linguistic and the indexical information in the speech signal better than the listeners with a lower voice recognition performance.[1] Additionally, similar to the analysis of both listening conditions together, a significant interaction was observed between day and voice recognition performance on day 2. This interaction is illustrated in Fig. 4, which plots the recognition accuracy of “old” words for the listeners with different voice recognition performance split out per day. As shown in Fig. 4, listeners with larger voice recognition performance were better at recognizing “old” words in general, except on the second day of the experiment. Finally, similar to the main analysis, listeners improved in their recognition of “old” words from day 1 to day 4.

The analysis of the noise listening condition showed a significant difference in performance for the words produced by the *familiar* and the *unfamiliar* talkers (see the right panel in Fig. 3 and the talker effect in Table 6), indicating that talker familiarity facilitated recognition memory in the noise listening condition. This familiarity benefit was not modulated by the lexical proficiency of the listeners or their voice recognition performance. Again,

Table 6 Estimates for the best fitting model for the d' measures of the “old” words in noise in the recognition memory task

Fixed effects	β	*SE*	*t*	*p*
Intercept	0.988	0.101	9.795	<0.001
Day 2	0.177	0.120	1.476	0.140
Day 3	**0.267**	**0.120**	**2.219**	**0.027**
Day 4	**0.356**	**0.120**	**2.966**	**0.003**
Talker	**−0.212**	**0.085**	**−2.498**	**0.012**
Random effects			*SD*	
Subject	intercept		0.192	

Significant factors and interactions are presented in bold

[1] To check whether higher accuracy of the listeners with a better voice recognition performance in recognizing “old” words can be explained by their larger working memory capacity in comparison to the listeners with a lower voice recognition performance, we ran an additional analysis in which working memory capacity was included in the model as a control variable. Working memory capacity was measured with a backward digit span task, completed by the listeners on day 2 (see Drozdova et al. (2017) for more information on the task). Inclusion of working memory capacity as a control variable did not change the result for the voice recognition performance factor. Moreover, voice recognition performance did not correlate with working memory capacity of the listeners. Hence, the observed results cannot be attributed to differences in working memory capacity between listeners.

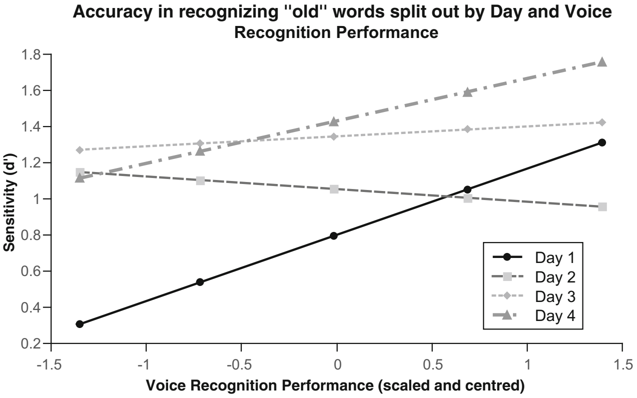

Fig. 4 Sensitivity *d′* in the recognition memory task in recognizing "old" words in the clear listening condition, split out by voice recognition performance and day

participants improved their performance from day 1 to day 4. In the noisy listening conditions, listeners did not profit from the knowledge they obtained about the different voices during the voice recognition training, as measured by the voice recognition performance. In both analyses, the final model only included subject as a random factor.

Reaction times

Reaction times, calculated from a word's offset, that were more than two standard deviations from the mean or were below zero were removed, which resulted in the deletion of about 4.5% of the data. The outliers were calculated separately for the clear and noise listening conditions. Figure 5 shows the log transformed reaction times measured from the word's offset for the correct identification of the "old" words when the words were presented in the clear (left panel) and in noise (right panel). Again, responses of the listeners to the words spoken by the *familiar* talkers are shown with the black dashed line with bullets and responses to the words spoken by the *unfamiliar* talkers are shown with the gray solid line with squares.

Log transformed offset reaction times were used as the dependent variable in a linear mixed effects model analysis. The initial model included talker (*familiar* or *unfamiliar*), day (as a categorical variable with day 1 as the reference), noise, voice recognition performance, lexical proficiency, and all possible interactions between them as fixed factors. Subject, item, talker number, and list were entered as random factors, with subject random slopes for noise, talker, and day, as listeners can differ in their sensitivity for noise and familiarity with the talker, and their improvement in recognition memory throughout the experiment and different talkers as familiar or unfamiliar could lead to different results. The estimates of the best-fitting model are presented in Table 7. The final best-fitting model for this analysis only included Subject, item, and talker number as random factors and by subject random slope for day, demonstrating that reaction times in correct recognition of words as "old" were not influenced by differences in participants' sensitivity to noise, or the combination of *familiar/unfamiliar* talkers.

Importantly, the analysis revealed a significant interaction between talker and lexical proficiency. Figure 6 illustrates this interaction by plotting the difference in reaction times for the words produced by *familiar* and *unfamiliar* talkers on different proficiency levels (the dots represent the mean difference in reaction times for each proficiency level) with the regression line. As shown in Fig. 6, the differences in reaction times to the words spoken by *familiar* and *unfamiliar* talkers were higher for listeners with a higher lexical proficiency than for listeners with a lower lexical proficiency, meaning that the familiar talker benefit was larger for more proficient listeners.

Moreover, there was a significant interaction between the factors talker and day on day 4, indicating that the difference between *familiar* and *unfamiliar* talkers increased on the

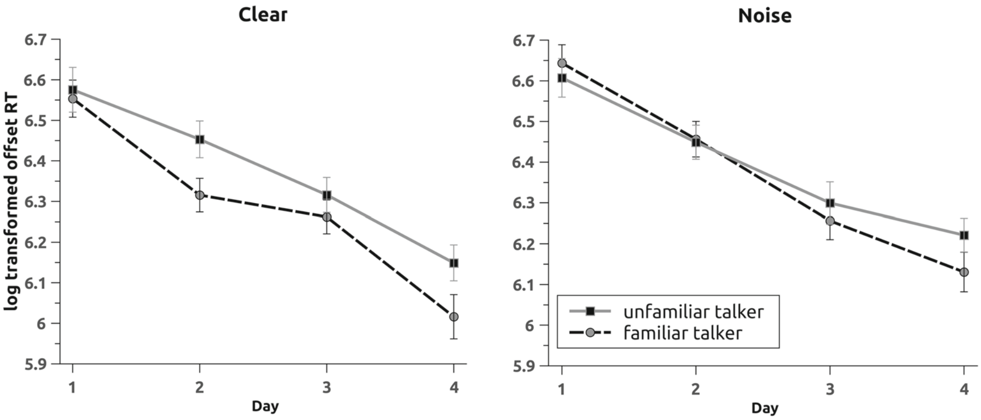

Fig. 5 Response times and standard errors for correct identification as "old" of the words spoken by the *familiar* (*black dashed line with bullets*) and *unfamiliar* talkers (*gray solid line with squares*), split out by listening condition (clear vs. noise)

Table 7 Estimates for the best fitting model for the reaction times of the hits for the "old" words in the recognition memory task

Fixed effects	*β*	*SE*	*t*	*p*
Intercept	6.602	0.074	88.85	<0.001
Day 2	**−0.235**	**0.071**	**−3.33**	**0.001**
Day 3	**−0.359**	**0.069**	**−5.16**	**<0.001**
Day 4	**−0.567**	**0.057**	**−9.90**	**<0.001**
Talker	−0.018	0.030	−0.27	0.787
Noise	**0.087**	**0.026**	**3.23**	**0.001**
Lexical Proficiency	0.003	0.050	0.06	0.951
Voice Recognition Performance	−0.023	0.029	−0.81	0.417
Day 2 x Talker	0.104	0.089	1.17	0.240
Day 3 x Talker	0.096	0.088	1.09	0.275
Day 4 x talker	**0.147**	**0.062**	**2.37**	**0.018**
Lexical proficiency x talker	**0.065**	**0.021**	**3.09**	**0.002**
Voice Recognition Performance x Talker	0.022	0.025	0.90	0.367
Voice recognition performance x noise	**0.060**	**0.023**	**2.61**	**0.009**
Noise x Talker	−0.040	0.040	−1.02	0.308
Noise x talker x voice recognition performance	**−0.072**	**0.034**	**−2.13**	**0.033**
Random effects			*SD*	
Item	intercept		0.182	
Talker Number	intercept		0.088	
Subject	intercept		0.284	
	Day 2		0.198	
	Day 3		0.164	
	Day 4		0.184	

Significant factors and interactions are presented in bold

last experimental day in comparison to the first experimental day. Talker also entered a significant three-way interaction with noise and voice recognition performance, indicating that the difference between words spoken by *familiar* and *unfamiliar* talkers was modulated by the voice recognition performance on a particular training day, depending on the listening condition. Similar to the accuracy analysis, the listeners demonstrated improvement from the first to the last experimental day: they became faster at giving correct responses to the "old" words (see also Fig. 5). Furthermore, participants were in general slower to react to words in noise than to words in the clear, and this difference was larger for

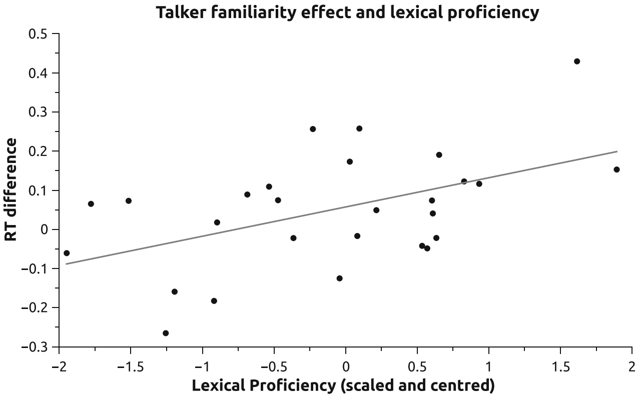

Fig. 6 The difference in reaction times (average RT for the words produced by *unfamiliar* talkers - average RT for the words produced by *familiar* talkers) in recognizing "old" words for different proficiency levels

the listeners with a better voice recognition performance on a given day (see interaction between voice recognition performance and noise). To investigate the differences between the clear and noise listening conditions and the effect of voice recognition performance on the talker familiarity benefit in more detail, separate analyses were carried out for the clear and noise listening conditions. The best-fitting model for the reaction times for the words in clear is presented in Table 8, and the best-fitting model for the words in noise is presented in Table 9.

The best-fitting model for the reaction times for the words in the clear and for the words in noise included random intercepts for subject, item, and talker number in the random structure. In the clear condition, the listeners significantly decreased their reaction times for the words from day 1 to day 4 (left panel of Fig. 5). Similar to the general analysis, a significant interaction between talker and lexical proficiency was found, indicating that in the clear listening condition, the difference between the words produced by *familiar* and *unfamiliar* talkers (faster reaction times to the words produced by *familiar* talkers) was only present for the listeners with a higher lexical proficiency.

The analysis of the words in the noise listening condition showed significantly faster reaction times for the words in noise as the experiment progressed (factor day; right panel of Fig. 5). No difference, however, was observed for the words produced by the *familiar* and *unfamiliar* talkers. Note that voice recognition performance did not come out as a significant effect in any of the separate analyses, although it entered a number of significant interactions in the analysis of clear and noise listening conditions together.

Table 8 Estimates for the best-fitting model for the reaction times of the hits for the "old" words in the recognition memory task in the clear listening condition

Fixed effects	*β*	*SE*	*t*	*p*
Intercept	6.555	0.071	92.14	<0.001
Day 2	**−0.194**	**0.054**	**−3.575**	**<0.001**
Day 3	**−0.282**	**0.053**	**−5.294**	**<0.001**
Day 4	**−0.500**	**0.044**	**−11.486**	**<0.001**
Talker	0.075	0.045	1.657	0.097
Lexical Proficiency	0.009	0.051	0.184	0.854
Lexical proficiency x talker	**0.075**	**0.030**	**2.509**	**0.012**
Random effects			*SD*	
Item	intercept		0.179	
Talker Number	intercept		0.102	
Subject	intercept		0.260	

Significant factors and interactions are presented in bold

Table 9 Estimates for the best-fitting model for the reaction times of the hits for the "old" words in the recognition memory task in the noise listening conditions

Fixed effects	*β*	*SE*	*t*	*p*
Intercept	6.673	0.065	102.2	<0.001
Day 2	**−0.192**	**0.052**	**−3.72**	**<0.001**
Day 3	**−0.362**	**0.051**	**−7.05**	**<0.001**
Day 4	**−0.499**	**0.042**	**−11.88**	**<0.001**
Random effects			*SD*	
Item	intercept		0.172	
Talker Number	intercept		0.089	
Subject	intercept		0.266	

Significant factors and interactions are presented in bold

To summarize, indexical information was used by non-native listeners in the recognition memory task: the listeners demonstrated a higher accuracy for the "old" words (measured with *d'*) when these words were produced by *familiar* talkers than when these items were produced by *unfamiliar* talkers, but this talker familiarity benefit only came out when noise was present. In the clear listening condition, another indexical effect was observed in the form of voice recognition performance. Listeners with higher voice recognition performance in the voice recognition training were also better at distinguishing "old" words from "new" words, except on day 2. This finding shows that listeners who showed improvement in voice recognition, could use the acquired knowledge to successfully identify "old" words. The talker familiarity benefit also revealed itself as shorter reaction times for the words produced by *familiar* talkers compared to words produced by *unfamiliar* talkers for listeners with higher lexical proficiency, but only in the clear listening condition. This finding demonstrates that non-native listeners should have a sufficient proficiency in their non-native language in order to benefit from talker familiarity at least in clear listening conditions.

Talker familiarity benefit in word identification

Before the analysis responses of the participants in the word identification task were coded as 1 if the answer was correct and 0 if the answer was incorrect. Obvious typing errors were corrected.

Figure 7 shows that the proportions of correctly identified words increased as listening conditions became easier. Importantly, both participant groups seemed to perform better on day 4 than on day 1 (the lines in the right panel are higher than the lines in the left panel), so also the listeners who did not receive any training on the target voice showed

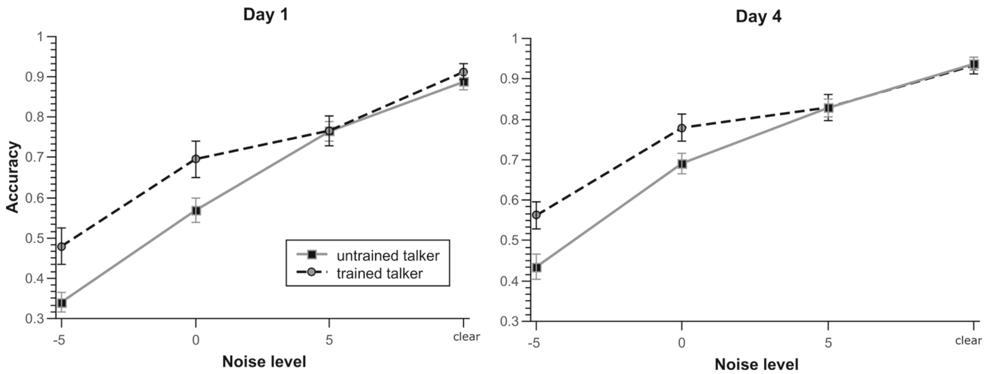

Fig. 7 Word identification accuracy with standard errors of the two listener groups for the four noise conditions. The *left panel* shows the results for the first experimental day; the *right panel* shows the results for the last experimental day. Responses of the listeners in the *trained* talker condition are presented with the *black dashed line with bullets*. Responses of the listeners in the *untrained* talker condition are presented with the *gray sold line with squares*

an increase in word identification accuracy (see the solid lines in Fig. 7). At the same time, the plots show that already on day 1 the word identification performance in the *trained* talker condition was higher than that of the *untrained* talker condition, which theoretically means that the participants in the *trained* talker condition could improve less than those in the *untrained* talker condition. We therefore calculated a measure of word identification improvement that takes into account a listener's maximum possible improvement which we refer to as "relative progress":

$$(a_2 - a_1)/(1 - a_1)$$

where a_1 is word identification performance (proportion of correct responses) of the participant on the first training day and a_2 is performance of the participant on the last training day.

To investigate whether the listeners from the *trained* talker condition demonstrated more progress, i.e., improved more, than the listeners from the *untrained* talker condition and whether this difference in improvement was modulated by the presence of noise, relative progress was used as the dependent variable in a linear mixed effects model analysis with SNR (-5, 0, 5, clear: with clear as a reference) and talker condition (*trained* talker versus an *untrained* talker) and their interaction as fixed factors. Additionally, lexical proficiency (i.e., the centered and scaled LexTALE score) was added as a fixed factor and in interaction with other factors. Subject was added as a random factor. The estimates from the best-fitting model are presented in Table 10. A significant effect of talker condition and/or in interaction with SNR or lexical proficiency would indicate a role of talker familiarity in word identification. However, as shown in Table 10, the analysis showed no such effect or interaction, indicating that the improvement in word identification performance was not different for the listeners who received the training in the target voice and those listeners who did not receive any training. To summarize, no effect of talker familiarity was observed in the word identification task.[2]

Discussion

The role of talker familiarity in recognition memory and word identification

The present study is the first study that investigated talker familiarity benefit in recognition memory and word identification for words spoken in a non-native language, and the role of background noise and non-native proficiency on this talker familiarity effect. After successfully learning to recognize four previously unfamiliar talkers over the course of four days, the familiar talker benefit in non-native listening was observed for recognition memory in line with other studies for native (Bradlow & Pisoni, 1999; Cooper & Bradlow, 2017; Goh, 2005; Goldinger, 1996; Luce & Lyons, 1998; Palmeri et al., 1993; Sheffert, 1998) and non-native listening (Trofimovich, 2005; Winters et al., 2013). At the same time, no talker familiarity benefit was observed during non-native word identification in

[2] Following Levi et al. (2011) and Nygaard and Pisoni (1998), a second analysis was conducted including voice recognition performance on the last training day as a potential predictor of relative progress. Possibly only listeners with a high voice recognition score on the final training day benefited from talker familiarity during word identification. In that case we would expect to find voice recognition performance influencing relative progress for the *trained* talker condition rather than the *untrained* talker condition (i.e., a higher relative progress for the listeners with a higher d' score in the *trained* talker condition). However, addition of the talker condition and voice recognition performance interaction did not significantly improve model fit (χ^2 (3)=1.282, p=0.734).

Table 10 Estimates for the best-fitting model of the word identification task

Fixed effects	*β*	*SE*	*t*	*p*
Intercept	0.292	0.085	3.437	<0.001
Lexical proficiency	**0.247**	**0.085**	**2.902**	**0.004**
SNR -5	−0.167	0.115	−1.461	0.144
SNR 0	−0.092	0.115	−0.797	0.425
SNR 5	−0.162	0.115	−1.413	0.158
SNR -5 x lexical proficiency	**−0.260**	**0.115**	**−2.260**	**0.024**
SNR 0 x lexical proficiency	**−0.291**	**0.115**	**−2.527**	**0.011**
SNR 5 x lexical proficiency	**−0.274**	**0.115**	**−2.375**	**0.018**
Random effects			*SD*	
Subject	intercept		0.143	

Significant factors and interactions are presented in bold

the clear and noisy listening conditions. Our results on the talker familiarity benefit in word identification, thus, do not correspond to the results observed for the native listeners in word identification (Nygaard & Pisoni, 1998, 1994), but are in line with the results found by Levi and colleagues who trained listeners to recognize talkers in an unfamiliar, although phonotactically related language. Several explanations for the absence of the talker familiarity benefit in non-native word identification in the present experiment can be offered.

Firstly, explanations for the absence of the talker familiarity benefit in the word identification task may be sought in differences in the design of the word identification task in the present study compared to that of others who investigated talker familiarity in native word identification. In our design, we used the same words, albeit different renditions of these words in the pre-and post-tests. Goldinger (1996) showed that the effects of indexical information emerge in a word-in-noise identification task even after a one-week delay. The match in indexical and linguistic information between the first and last training day in our experiment could, therefore, in principle, have led to an improvement in performance for both groups (i.e., the *trained* talker condition and *untrained* talker condition). If this explanation is correct, it would mean that even the small amount of exposure to the target voice on day 1 was sufficient to generate familiarity with the voice, so that the additional exposure to the voice on the next days generated little additional benefit (see also Newman and Evers (2007)). This explanation, however, contradicts the finding in the recognition memory task, where larger differences in the reaction times to words produced by *familiar* talkers and *unfamiliar* talkers were observed on the last experimental day than on the first experimental day, suggesting that the talker familiarity benefit increases over time as familiarity with the talker increases. To further probe this hypothesis, future research should include novel words in the post-test phase to test whether talker familiarity effect generalizes to words not present on the first experimental day.

Furthermore, in the present study the voice of a single talker was used in the word identification task, while multiple familiar and unfamiliar talkers were used in the previous studies on talker familiarity effects in word identification in noise (Nygaard & Pisoni, 1998, 1994; Yonan & Sommers, 2000). Changes in voice from trial to trial were shown to slow down lexical processing in previous studies with native (Mullennix et al., 1989; Ryalls & Pisoni, 1997; Sommers et al., 1994) and non-native listeners (Bradlow & Pisoni, 1999; Tamati & Pisoni, 2014). Given that in the present study the voice was kept constant in the word identification task, it is possible that the speech processing and word recognition was not slow enough for indexical information to be accessed. At the same time, since listening occurred in noise and in a non-native language, this explanation does not seem likely.

Instead, the absence of a talker familiarity benefit in the word identification task could potentially be explained by the differences between native and non-native listeners, either related to the task demands or the type of information encoded during voice recognition training. Non-native listeners have greater difficulty processing speech than native listeners, especially in the presence of noise (Borghini & Hazan, 2018; Broersma, 2012; Garcia Lecumberri et al., 2010; Scharenborg et al., 2018; Weber & Cutler, 2004). It is possible that the word identification task was too hard for the listeners, which prevented them from benefiting from talker familiarity on the last experimental day. However, if this explanation is correct, we would have expected to find effects of noise level and lexical proficiency in the word identification tasks, with the listeners benefiting from the familiarity with the voice of the talker only when they had a certain non-native proficiency and only at certain noise levels. This is, however, not what we observed. So, greater task difficulty for the non-native compared to the native listeners as such is an unlikely explanation for the absence of a familiar talker effect in non-native word identification. A more likely explanation has to do with the acoustic and linguistics information encoded by the native and non-native listener. Mattys et al. (2010) demonstrated that, compared to native listeners, non-native speakers pay less attention to lexical information and relatively more attention to acoustic detail when processing speech. At the same time, Levi (2014) hypothesized that when listeners learn to recognize novel talkers in their native language, they not only attend to the voices but also automatically recognize and encode the words they hear. Not recognizing the words appeared to block the emergence of the familiar talker benefit as shown by the lack of the effect for the English listeners who were trained on German speech. The encoding of lexical information, on the other hand, allows the listener to perform well in the subsequent word identification task. Possibly, when voice recognition training is performed in a non-native language, the words are not processed deeply enough to ensure encoding of the linguistic and indexical information, and, as a result, a talker familiarity benefit does not emerge in a task which requires lexical access such as the word identification task.

In the same vein, one final possible explanation is related to the differences between the recognition memory and word identification tasks. Although a talker familiarity benefit was previously observed for native listeners in a word identification in noise task, mixed results were obtained for tasks involving lexical access in general. This is in contrast with tasks not requiring lexical access (e.g., recognition memory task), where the effects of indexical information were consistently observed. The discrepancies in the emergence of the effect of indexical information in these two different types of tasks have been previously observed by Luce and Lyons (1998), Kittredge et al. (2006), and Lee and Zhang (2015), who found the effect of indexical information in a recognition memory task and repetition priming, but not in a lexical decision task and during semantic priming. While the recognition memory task requires minimal contact with the mental lexicon and does not abstract away indexical information in spoken words (Cooper & Bradlow, 2017), the word identification task focuses on sound and lexical processing which does abstract from voice-specific characteristics (Cutler et al., 2010a, b). Indeed, even listeners not familiar with a language demonstrate effects of indexical information in recognition memory (Winters et al., 2013) but not in word identification (Levi et al., 2011).

To summarize, our results suggest that the emergence of the talker familiarity benefit in both native and non-native listening is dependent on the task, and consequently on the specific process involved in speech processing. Further, the emergence of the talker familiarity benefit seems to depend on the type of information (lexical or acoustic) encoded during the talker training, which is potentially different in native and non-native listening (Levi et al., 2011).

The effect of background noise and lexical proficiency on the talker familiarity effect

The second research question in the present study addressed the role of background noise and lexical proficiency of the listeners in the emergence of the talker familiarity benefit. In line with numerous studies on non-native speech comprehension in noise (e.g., Garcia Lecumberri et al., 2010; Scharenborg et al., 2018), the presence of background noise had a negative effect on speech comprehension. The results showed that participants were slower to react to words in noise than to words in the clear in the recognition memory task, while fewer words were recognized in worse listening conditions in the word identification task. According to the *time-course hypothesis* (McLennan & Luce, 2005), a larger talker familiarity benefit could be expected to occur for words in noise than in the clear in the recognition memory task as the effects of indexical information emerge relatively late in processing.

Our findings, however, do not fully support the *time-course hypothesis*. Although the presence of noise in the stimuli did influence the emergence of the talker familiarity benefit, its effect differed depending on whether accuracy or reaction times were measured. When accuracy was analyzed, the difference in recognition of the words produced by *familiar* and *unfamiliar* talkers was observed for the words in noise but not for the words in the clear which is in line with the *time-course hypothesis*. Note, however, that the combined analysis of the words in the clear and in noise revealed a significant effect of talker familiarity which was not modulated by listening conditions. The distinction between the two listening conditions seems to depend on a different trade-off between the effects of talker and voice recognition performance. To investigate this possibility, we used the BIC (Bayes Information Criterion) to select the best model in our data analyses. Comparing models on accuracy in the noise condition, a model with talker was clearly superior to a model with voice recognition performance (BIC 622.42 vs. 610.41, both models including the interaction effect with day; a lower BIC value is better). However, in the clear condition, the two effects competed, the model with the voice recognition performance being only slightly better than the model with the talker effect (BIC 591.42 vs. 593.58, both models including the interaction effect with day). Voice recognition performance overshadowed the talker effect. This suggests that also in the clear listening condition, listeners were more accurate at reacting to the words produced by *familiar* talkers than by *unfamiliar* talkers.

Moreover, in the clear listening condition, the listeners who were better at voice recognition during the voice recognition training demonstrated a more accurate recognition of the "old" words than the listeners who were worse at voice recognition, suggesting that these listeners could establish the connection between the linguistic and the indexical information in the speech signal better than the listeners with lower voice recognition performance. Both the difference in recognition of the words produced by *familiar* and *unfamiliar* talkers (talker effect) and voice recognition performance (measured for each individual listener) evidence the role of indexical information in recognition memory. As we know from previous research (Brouwer & Bradlow, 2011, 2016; Hintz & Scharenborg, 2016), listeners are less successful in extracting information from the speech signal and need more time to use these information sources for speech comprehension when the speech signal is noisy. So, it is plausible that individual differences between the listeners (the voice learning performance effect) fade away and that only a global talker effect remains when the items are presented in noise. This expectation is corroborated by the fact that the voice learning performance effect is absent in noise but not in the clear listening condition.

In the reaction time analysis, the talker familiarity benefit only emerged for words in the clear condition and only for the listeners with a higher lexical proficiency. These findings are not in agreement with the *time-course hypothesis*. The role of lexical proficiency fits in with the above proposed explanation for the lack of a talker familiarity benefit in word identification: non-native listeners with a higher proficiency in the non-native language have better representations of the non-native sounds and words and consequently are able to process the words more deeply which would consequently lead to the emergence of the talker familiarity effect. This would also be in line with a higher sensitivity to acoustic patterns in the speech signal by listeners with a higher lexical proficiency, who are, as a consequence, better at encoding indexical information and consequently better at discriminating *familiar* and *unfamiliar* talkers. A similar explanation was offered by Trofimovich (2008) who suggested that more experienced non-native listeners are likely to be better at encoding context-specific phonological information from non-native words, and thus have an improved encoding of indexical information and discrimination of *familiar* and *unfamiliar* talkers.

A possible explanation for the different effects of talker familiarity on reaction times and accuracy when listening in background noise could potentially be connected to the fact that, as suggested by MacLeod and Nelson (1984), accuracy and reaction times measure different aspects of memory. While accuracy measures the sufficiency of encoding for retrieval, reaction times measure the number of steps during the retrieval before a (correct) decision is made, and, therefore, refers to the process. Further research is needed to answer the question of at what stage of speech processing (during or after the retrieval) indexical information is accessed and what the role is of listening conditions.

Another explanation for the task-dependency of the emergence of the talker familiarity benefit was provided by Goh (2005), who found no reliable differences in reaction times between the words produced by *familiar* and *unfamiliar* talkers, but only found them in accuracy measures. He noted that previous studies observing the effects of indexical information in reaction times (Goldinger, 1996; Luce & Lyons, 1998) included switches between male and female talkers in their manipulations of voice changes, whereas in the study by Goh (2005) and in the present study only male talkers were used. Goh (2005) hypothesized that differences in reaction times between the words produced by familiar and unfamiliar talkers can become more pronounced when changes from talker to talker are made more distinctive as in the studies using talkers of different genders.

Interestingly, the familiar talker benefit in the recognition task was observed in the accuracy measures for the words in noise even though the voice recognition training was always conducted in the clear. In previous studies on the effects of indexical information in background noise (Goldinger, 1996; Nijveld et al., 2015), words in noise were used in both the exposure and test phases of the recognition memory task to avoid training-test format changes (Goldinger, 1996; Schacter & Church, 1992). The results of the present study demonstrate that talker familiarity in recognition memory generalizes to other listening conditions.

The role of talker familiarity in speech processing

The demonstration of the importance of talker-related information in speech processing in the 1990s (Mullennix et al., 1989; Nygaard et al., 1994, 1998; Palmeri et al., 1993) led to the emergence of exemplar-based theories of spoken word recognition (Goldinger, 1998; Pierrehumbert, 2001). These theories state that upon hearing a new word, detailed episodic information about this word is stored in the mental lexicon of the listeners. During speech comprehension, the listener compares the incoming acoustic information with the stored detailed representation. Exemplar-based theories challenged the standard abstractionist view of speech perception, which claimed that listeners map the words they hear onto abstract representations at the prelexical and lexical processing levels, while indexical information is discarded as irrelevant (see Pisoni (1997)). Although the present study confirmed that indexical information is stored in the memory of non-native listeners and is accessed during a recognition memory task, we found no evidence for the storage of the indexical information together with the lexical representation of a word as no talker familiarity effect was observed in the word identification task, which required lexical access.

Different studies (see, e.g., Cutler (2010a) and McLennan and Luce (2005)) expressed the need for a hybrid model of speech perception activating and exploiting both abstract representations and more specific form-based representations. Several attempts have been made in formulating such a hybrid theory (e.g., Cutler (2010a), Goldinger (2007), Kleinschmidt and Jaeger (2015), Luce et al. (2003), and McQueen et al. (2006)). For instance, theories implying Bayesian inference argue that listeners make and update predictions about the speech signal based on the available evidence (Kleinschmidt and Jaeger, 2015; Norris & McQueen, 2008, 2016). In this framework (e.g., Kleinschmidt and Jaeger (2015)), voice recognition and the familiar talker benefit can be explained by listeners creating a talker-specific generative model on the basis of talker-specific mappings of acoustic cues to phonetic categories. Listeners are able to recognize a familiar situation (familiar talker) and take advantage of this familiarity. At the same time, theories of this type imply that each successive input is used to update the belief of the listeners about the likelihood of a certain event occurring (Pufahl & Samuel, 2014), which could theoretically mean larger effects of talker-specific information for talkers to whom the listeners had more exposure. This is however not what we observe in our word identification task with non-native listeners. We did not find an additional familiarity advantage despite the listeners having had extensive training on the voice of the talker.

The results of the present study agree with another type of hybrid theories, namely, weak abstractionist hybrid theories (Cutler, 2010a, b). The proponents of these less strong abstractionist theories argue that although indexical effects show that indexical information is stored in the memory of the listeners, there is no evidence that this information is stored in the mental lexicon. According to these theories indexical information can be either stored prelexically, i.e., facilitating recognition of words containing prelexical perceptual units produced by the same or familiar talker or in an episodic memory system, separate from but linked to a linguistically abstract lexicon (Cutler et al., 2010a, b; Jesse et al., 2007).

Conclusions

The present study demonstrated that non-native listeners store indexical information from the speech signal in their memory. This indexical information is however not accessed at all times during speech processing. The facilitatory effect of indexical information is dependent on the specific stage of the speech recognition process, the listening conditions, proficiency in the non-native language, the degree of familiarity with the talker, and, potentially, the type of information (semantic or acoustic) encoded during talker familiarization. Specifically, the facilitatory effect of indexical information was only observed during a task involving recognition memory but not a task which involved lexical access, i.e., word identification. This clear distinction in the emergence of talker familiarity effects in the tasks with and without lexical access, together with results from other recent studies (Jesse et al., 2007; Kittredge et al., 2006; Lee & Zhang, 2015), seems to suggest that indexical information is not stored as an integral part of the lexical representations, which is in line with theories suggesting the existence of an episodic memory system, distinct from the mental lexicon but linked to a linguistically abstract lexical or prelexical level, where indexical information is stored (Cutler, 2010a).

Acknowledgements This research is supported by a Vidi-grant from the Netherlands Organization for Scientific Research (NWO; grant number 276-89-003) awarded to Odette Scharenborg. Odette Scharenborg is now at the Multimedia Computing Group, Delft University of Technology, the Netherlands.

Appendix

Table 11 Words used in the experiment with their corresponding frequencies

Word	Frequency (per million)	Word	Frequency
Access	56.23	Kite	7.76
Ache	2.23	Magic	67.61
Bacon	21.88	Mat	5.75
Basket	14.45	Money	691.83
Beach	52.48	Muffin	10
Bean	11.48	Neck	44.67
Beast	12.88	Niece	5.01
Bed	131.83	Nut	9.77
Body	151.36	Pace	39.81
Bonus	46.77	Pan	39.81
Book	162.18	Path	30.20
Bottom	112.20	Peak	18.20
Bunny	7.76	Photo	26.30
Canvas	8.91	Picnic	10.97
Cave	15.49	Pig	32.36
City	251.19	Pumpkin	6.17
Coffee	44.67	Pun	2.57
Day	977.24	Sack	10
Diet	18.20	Skeptic	2.19
Duck	33.88	Shop	112.20
Event	72.44	Shout	26.92
Game	346.74	Sketch	7.08
Giant	43.65	Sofa	16.60
Heat	57.54	Speed	81.28
Heaven	39.81	Team	467.74
Human	114.82	Ticket	32.36
Husband	87.10	Tip	35.48
Income	43.65	Topic	6.76
Index	4.68	Vision	36.31
King	123.02	Zombie	2.69

Word identification task

Table 12 Voice recognition training (familiarization phase)

Word	Frequency (per million)	Word	Frequency (per million)
Baby	4.07	Oven	30.90
Object	194.98	Atom	34.67
Sip	131.83	Woman	33.11
Socket	3.80	Hit	4.17
Case	269.15	Gang	4.17
Music	204.17	Peach	2.14
Page	24.55	Campus	2.14
Paw	4.90	Tin	3.63
Soda	204.17	Ocean	23.99
System	33.11	Map	181.97
Box	223.87	Status	22.39
Heap	29.51	Fun	165.96

Table 13 Voice recognition training (feedback and test phases)

Word	Frequency (per million)	Word	Frequency (per million)
Accent	14.45	Movie	41.69
Action	123.03	Name	407.38
Aspect	17.38	Nation	57.54
Baggage	3.09	Nest	25.12
Bat	19.95	Note	41.69
Bath	44.67	Nothing	436.52
Beam	9.33	Nun	2.95
Bee	15.49	Oak	16.98
Biscuit	12.59	Office	134.90
Bucket	16.22	Onion	19.06
Bug	11.75	Option	40.74
Bun	3.39	Pack	41.69
Bus	54.95	Package	22.39
Bush	19.50	Pants	19.50
Cabbage	11.75	Pasta	15.49
Cake	64.57	Peace	54.95
Caution	6.46	Pen	23.44
Comet	3.72	Penguin	7.59
Contact	52.48	Physics	9.12
Contest	22.38	Pick	144.54
Cook	107.15	Pin	18.20
Copy	23.99	Pinch	10.23
Cotton	12.59	Pizza	17.78
Cousin	18.20	Pocket	42.66
Cuff	1.86	Poison	8.32
Cupcake	1.95	Poppy	9.55
Cut	223.87	Pub	51.29
Damage	64.57	Pudding	27.54
Deed	3.31	Question	398.11
Demon	3.39	Sanction	1.38
Diamond	24.55	Sandwich	17.78
Disco	10.23	Sausage	16.98
Dish	87.10	Sentence	21.88
Dust	24.55	Shampoo	2.57
End	588.84	Shoe	14.13
Fame	15.14	Size	123.03
Fate	16.60	Smoothie	1.86
Feast	12.59	Sob	1.02
Finance	25.70	Stag	7.59
Finish	114.82	Stocking	2.34
Function	11.48	Subject	54.95
Habit	10.96	Suit	38.90
Hedge	9.12	Sum	17.38
Hen	9.12	Sun	102.33
Hip	23.44	Tab	1.91
Hook	16.22	Tan	11.48
Hunt	48.98	Tango	7.59
Idea	363.08	Tap	21.38
Image	43.65	Tennis	27.54
Jacket	19.50	Tent	16.22
Jam	21.88	Test	114.82
Jaw	7.59	Tissue	9.33
Job	398.11	Toe	12.02
Kick	81.28	Top	436.52
Kid	60.26	Touch	125.89
Kin	3.47	Union	75.86
Kitchen	158.49	Wand	3.47
Mass	33.88	Wedding	89.13
Meat	63.10	Weekend	93.33
Message	87.10	Wheat	8.91
Minute	169.82	Wig	6.61
Mistake	61.66	window	69.18
Moment	371.54	Wing	28.84
Mountain	43.65	Witness	19.06

Table 14 New words from the old/new task

Word	Frequency (per million)	Word	Frequency (per million)
Absence	7.76	Kiss	44.67
Acid	10.23	Knob	3.162
Amount	117.49	Man	724.44
Attack	79.43	Mason	8.51
Attempt	39.81	Meaning	33.11
Auction	147.91	Menu	22.91
Autumn	25.70	Mix	50.12
Axis	1.78	Monkey	23.99
Back	1778.28	Mop	2.88
Beak	6.76	Motion	21.38
Bit	1258.93	Mud	23.44
Bite	28.18	Net	33.88
Boot	26.92	Oath	3.47
Budget	154.88	Outcome	19.05
Business	295.12	Outfit	13.49
Button	28.84	Pad	9.33
Cactus	2.24	Panda	5.37
Camping	8.91	Pass	138.04
Cape	8.91	Paste	8.32
Captain	70.79	Peacock	4.68
Cash	89.13	Peanut	4.47
Casket	1.20	Penny	30.90
Cast	38.02	Pit	15.85
Cave	15.49	Pity	12.59
Champion	64.57	Pod	23.99
Chat	39.81	pot	57.54
Check	125.90	Potion	1.78
Chicken	66.07	Pup	5.50
Chimney	7.94	Quote	14.13
Concept	17.38	Saga	3.39
Cowboy	11.48	Saint	11.75
Cup	123.03	Saw	251.19
Cushion	17.78	Science	63.10
Debate	83.18	Seat	60.26
Deck	12.59	Session	21.38
Defence	77.63	Shed	23.99
Dip	16.60	Shock	42.66
Disease	30.90	Sight	43.65
Dispute	11.22	Snake	22.91
Distance	43.65	Soap	12.59
Donkey	9.77	Song	125.89
Dot	16.60	Squad	24.55
Dozen	12.59	Stain	3.39
Duet	2.95	Statement	56.23
Edge	63.10	Station	66.07
Egg	63.10	Stick	104.71
Evening	131.83	Stomach	19.06
Fashion	42.66	Student	40.74
Fight	109.65	Subway	1.78
Fitness	9.33	Success	83.18
Focus	60.26	Swan	9.55
Font	1.20	Tame	3.89
Fuss	8.91	Taxes	23.44
Gaze	3.55	Tension	16.22
Goodness	53.70	Thumb	9.33
Haven	7.59	Tick	14.13
Honey	39.81	Tone	15.14
Hood	8.91	Tongue	22.91
Hop	20.42	Tube	19.05
Hostess	1.51	Type	85.11
Insect	7.08	Venue	16.22
Jet	16.22	Visit	87.10
Joke	38.90	Wish	112.20
Kidney	8.71	Witch	11.48

References

Abercombie, D. (1967). *Elements of general phonetics*. London: Aldine Pub. Company.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Boersma, P., & Weenink, D. (2009). Praat: doing phonetics by computer (version 5.1. 05)[computer program]. retrieved may 1, 2009.

Borghini, G., & Hazan, V. (2018). Listening effort during sentence processing is increased for non-native listeners: A pupillometry study. *Frontiers in Neuroscience*, *12*, 152.

Bradlow, A. R., & Pisoni, D. B. (1999). Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors. *The Journal of the Acoustical Society of America*, *106*(4), 2074–2085.

Bregman, M. R., & Creel, S. C. (2014). Gradient language dominance affects talker learning. *Cognition*, *130*(1), 85–95.

Broersma, M. (2012). Increased lexical activation and reduced competition in second-language listening. *Language and Cognitive Processes*, *27*(7-8), 1205–1224.

Brouwer, S., & Bradlow, A. R. (2011). The influence of noise on phonological competition during spoken word recognition. In *Proceedings of the international congress of phonetic sciences. International Congress of Phonetic Sciences*, (Vol. 2011, p. 364).

Brouwer, S., & Bradlow, A. R. (2016). The temporal dynamics of spoken word recognition in adverse listening conditions. *Journal of Psycholinguistic Research*, *45*(5), 1151–1160.

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America*, *116*(6), 3647–3658.

Cooper, A., & Bradlow, A. R. (2017). Talker and background noise specificity in spoken word recognition memory. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *8*(1):29, pp. 1–15. https://doi.org/10.5334/labphon.99

Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2008). Heeding the voice of experience: The role of talker variation in lexical access. *Cognition*, *106*(2), 633–664.

Cutler, A. (2010a). Abstraction-based efficiency in the lexicon. *Laboratory Phonology*, *1*(2), 301–318.

Cutler, A., Eisner, F., McQueen, J. M., & Norris, D. (2010b). How abstract phonemic categories are necessary for coping with speaker-related variation. *Laboratory Phonology*, *10*, 91–111.

Dahan, D., Drucker, S. J., & Scarborough, R. A. (2008). Talker adaptation in speech perception: Adjusting the signal or there presentations?. *Cognition*, *108*(3), 710–718.

Drozdova, P., Van Hout, R., & Scharenborg, O. (2016). Lexically-guided perceptual learning in non-native listening. *Bilingualism: Language and Cognition*, *19*(5), 914–920.

Drozdova, P., van Hout, R., & Scharenborg, O. (2017). L2 voice recognition: The role of speaker-, listener-, and stimulus-related factors. *The Journal of the Acoustical Society of America*, *142*(5), 3058–3068.

Dufour, S., & Nguyen, N. (2014). Access to talker-specific representations is dependent on word frequency. *Journal of Cognitive Psychology*, *26*(3), 256–262.

Dufour, S., Bolger, D., Massol, S., Holcomb, P. J., & Grainger, J. (2017). On the locus of talker-specificity effects in spoken word recognition: an ERP study with dichotic priming. *Language, Cognition and Neuroscience*, *32*(10), 1273–1289.

Garcia Lecumberri, M. L., Cooke, M., & Cutler, A. (2010). Non-native speech perception in adverse conditions: A review. *Speech Communication*, *52*(11), 864–886.

Goh, W. D. (2005). Talker variability and recognition memory: instance-specific and voice-specific effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 40.

Goldinger, S. D. (1996). Words and voices: episodic traces in spoken word identification and recognition memory. *Journal of experimental psychology: Learning, memory, and cognition*, *22*(5), 1166.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*(2), 251.

Goldinger, S. D. (2007). A complementary-systems approach to abstract and episodic speech perception. In *Proceedings of the 16th International Congress of Phonetic Sciences*, (pp. 49–54). Saarbrúcken.

Hintz, F., & Scharenborg, O. (2016). The effect of background noise on the activation of phonological and semantic information during spoken-word recognition.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards log it mixed models. *Journal of Memory and Language*, *59*(4), 434–446.

Jesse, A., McQueen, J. M., & Page, M. (2007). The locus of talker-specific effects in spoken-word recognition. *16th international congress of phonetic sciences (ICPhS 2007)*, 1921–1924.

Kittredge, A., Davis, L., & Blumstein, S. E. (2006). Effects of nonlinguistic auditory variations on lexical processing in Broca's aphasics. *Brain and Language*, *97*(1), 25–40.

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, *122*(2), 148.

Lee, C. Y., & Zhang, Y. (2015). Processing speaker variability in repetition and semantic/associative priming. *Journal of Psycholinguistic Research*, *44*(3), 237–250.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, *44*(2), 325–343.

Levi, S. V., Winters, S. J., & Pisoni, D. B. (2011). Effects of cross-language voice training on speech perception: Whose familiar voices are more intelligible? *The Journal of the Acoustical Society of America*, *130*(6), 4053–4062.

Levi, S. V. (2014). Individual differences in learning talker categories: The role of working memory. *Phonetica*, *71*(3), 201–226.

Luce, P. A., & Lyons, E. A. (1998). Specificity of memory representations for spoken words. *Memory & Cognition*, *26*(4), 708–715.

Luce, P. A., McLennan, C. T., & Chance-Luce, J. (2003). Abstractness and specificity in spoken word recognition: Indexical and allophonic variability in long-term repetition priming. In B. J, & M. C (Eds.) *Rethinking implicit memory*, (pp. 197–214). Oxford: Oxford University Press.

MacLeod, C. M., & Nelson, T. O. (1984). Response latency and response accuracy as measures of memory. *Acta Psychologica*, *57*(3), 215–235.

Macmillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group data: Estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin*, *98*(1), 185.

Maibauer, A. M., Markis, T. A., Newell, J., & McLennan, C. T. (2014). Famous talker effects in spoken word recognition. *Attention, Perception, and Psychophysics*, *76*(1), 11–18.

Mattys, S. L., & Liss, J. M. (2008). On building models of spoken-word recognition: When there is as much to learn from natural "oddities" as artificial normality. *Attention, Perception, and Psychophysics*, *70*(7), 1235–1242.

Mattys, S. L., Carroll, L. M., Li, C. K., & Chan, S. L. (2010). Effects of energetic and informational masking on speech segmentation by native and non-native speakers. *Speech Communication*, *52*(11), 887–899.

Mayo, L. H., Florentine, M., & Buus, S. (1997). Age of second-language acquisition and perception of speech in noise. *Journal of Speech, Language, and Hearing Research*, *40*(3), 686–693.

McLennan, C. T., & González, J. (2012). Examining talker effects in the perception of native-and foreign-accented speech. *Attention, Perception, and Psychophysics*, *74*(5), 824–830.

McLennan, C. T., & Luce, P. A. (2005). Examining the time course of indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(2), 306.

McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science*, *30*(6), 1113–1126.

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *The Journal of the Acoustical Society of America*, *85*(1), 365–378.

Newman, R. S., & Evers, S. (2007). The effect of talker familiarity on stream segregation. *Journal of Phonetics*, *35*(1), 85–103.

Nijveld, A., ten Bosch, L., & Ernestus, M. (2015). Exemplar effects arise in a lexical decision task but only under adverse listening conditions. In *Proceedings of the 18th international congress of phonetic sciences (ICPhS 2015)*. Glasgow: University of Glasgow.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*(2), 204–238.

Norris, D., & McQueen, J. M. (2008). Shortlist B: a Bayesian model of continuous speech recognition. *Psychological Review*, *115*(2), 357.

Norris, D., McQueen, J. M., & Cutler, A. (2016). Prediction, Bayesian inference and feedback in speech recognition. *Language, Cognition and Neuroscience*, *31*(1), 4–18.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, *5*(1), 42–46.

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Attention, Perception, and Psychophysics*, *60*(3), 355–376.

Nygaard, L. C., Sidaras, S. K., & Alexander, J. E. (2008). Time course of talker-specific learning in spoken word recognition. *The Journal of the Acoustical Society of America*, *124*(4), 2459–2459.

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(2), 309.

Papesh, M. H., Goldinger, S. D., & Hout, M. C. (2016). Eye movements reveal fast, voice-specific priming. *Journal of Experimental Psychology: General*, *145*(3), 314.

Perrachione, T. K., & Wong, P. C. (2007). Learning to recognize speakers of a non-native language: Implications for the functional organization of human auditory cortex. *Neuropsychologia*, *45*(8), 1899–1910.

Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition and contrast, frequency and the emergence of linguistic structure. In Bybee, J., & Hopper, P. (Eds.), (pp. 137–57). Amsterdam: John Benjamins.

Pisoni, D. B. (1997). Some thoughts on "normalization" in speech perception. In K. Johnson & J. W. Mullennix (Eds.), *Talker Variability in Speech Processing* (pp. 9–32). San Diego: Academic Press.

Pufahl, A., & Samuel, A. G. (2014). How lexical is the lexicon? evidence for integrated auditory memory representations. *Cognitive Psychology*, *70*, 1–30.

Reinisch, E., Weber, A., & Mitterer, H. (2013). Listeners retune phoneme categories across languages. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(1), 75–86.

Rogers, C. L., Lister, J. J., Febo, D. M., Besing, J. M., & Abrams, H. B. (2006). Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. *Applied Psycholinguistics*, *27*(3), 465–485.

Ryalls, B. O., & Pisoni, D. B. (1997). The effect of talker variability on word recognition in preschool children. *Developmental Psychology*, *33*(3), 441.

Schacter, D. L., & Church, B. A. (1992). Auditory priming: implicit and explicit memory for words and voices. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(5), 915.

Scharenborg, O., Coumans, J. M., & van Hout, R. (2018). The effect of background noise on the word activation process in nonnative spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(2), 233.

Sheffert, S. M. (1998). Voice-specificity effects on auditory word priming. *Memory & Cognition*, *26*(3), 591–598.

Sommers, M. S., Nygaard, L. C., & Pisoni, D. B. (1994). Stimulus variability and spoken word recognition. I. effects of variability in speaking rate and overall amplitude. *The Journal of the Acoustical Society of America*, *96*(3), 1314–1324.

Tamati, T. N., & Pisoni, D. B. (2014). Non-native listeners' recognition of high-variability speech using presto. *Journal of the American Academy of Audiology*, *25*(9), 869–892.

Theodore, R. M., Blumstein, S. E., & Luthra, S. (2015). Attention modulates specificity effects in spoken word recognition: Challenges to the time-course hypothesis. *Attention, Perception, and Psychophysics*, *77*(5), 1674–1684.

Trofimovich, P. (2005). Spoken-word processing in native and second languages: An investigation of auditory word priming. *Applied Psycholinguistics*, *26*(4), 479–504.

Trofimovich, P. (2008). What do second language listeners know about spoken words? Effects of experience and attention in spoken word processing. *Journal of Psycholinguistic Research*, *37*(5), 309–329.

Trude, A. M., & Brown-Schmidt, S. (2012). Talker-specific perceptual adaptation during online speech perception. *Language and Cognitive Processes*, *27*(7-8), 979–1001.

Tuft, S. E., McLennan, C. T., & Krestar, M. L. (2018). Hearing taboo words can result in early talker effects in word recognition for female listeners. *Quarterly Journal of Experimental Psychology*, *71*(2), 435–448. https://doi.org/10.1080/17470218.2016.1253757

Van Heuven, W. J., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190.

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, *50*(1), 1–25.

Winters, S., Lichtman, K., & Weber, S. (2013). The role of linguistic knowledge in the encoding of words and voices in memory. In *Second language research forum, Ames, Iowa*.

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*(1), 88.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.