THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY
2011, 64 (7), 1442–1456

Psychology Press
Taylor & Francis Group

When less is heard than meets the ear: Change deafness in a telephone conversation

Kimberly M. Fenn[1], Hadas Shintel[2,3], Alexandra S. Atkins[4], Jeremy I. Skipper[5], Veronica C. Bond[2,3], and Howard C. Nusbaum[2,3]

[1]Department of Psychology, Michigan State University, East Lansing, MI, USA
[2]Department of Psychology, University of Chicago, Chicago, IL, USA
[3]Center for Cognitive and Social Neuroscience, University of Chicago, Chicago, IL, USA
[4]Department of Psychology, Duke University, Durham, NC, USA
[5]Department of Psychology, Hamilton College, Clinton, NY, USA

During a conversation, we hear the sound of the talker as well as the intended message. Traditional models of speech perception posit that acoustic details of a talker's voice are not encoded with the message whereas more recent models propose that talker identity is automatically encoded. When shadowing speech, listeners often fail to detect a change in talker identity. The present study was designed to investigate whether talker changes would be detected when listeners are actively engaged in a normal conversation, and visual information about the speaker is absent. Participants were called on the phone, and during the conversation the experimenter was surreptitiously replaced by another talker. Participants rarely noticed the change. However, when explicitly monitoring for a change, detection increased. Voice memory tests suggested that participants remembered only coarse information about both voices, rather than fine details. This suggests that although listeners are capable of change detection, voice information is not continuously monitored at a fine-grain level of acoustic representation during natural conversation and is not automatically encoded. Conversational expectations may shape the way we direct attention to voice characteristics and perceive differences in voice.

Keywords: Change detection; Speech perception; Voice memory; Auditory communication.

Research on visual scene perception has found that people are often unaware of large visual changes to a two-dimensional or three-dimensional scene (Grimes, 1996; Hollingworth, 2004; Hollingworth & Henderson, 2002; Mitroff & Simons, 2002; O'Regan, Rensink, & Clark, 1999; Scholl, 2000; Simons, 1996; Simons, Chabris, Schnur, & Levin, 2002; Simons, Franconeri, & Reimer, 2000; for reviews see Rensink, 2002; Simons, 2000; Simons & Levin, 1997; Simons &

Correspondence should be addressed to Kimberly M. Fenn, Department of Psychology, Michigan State University, East Lansing, MI 48824, USA. E-mail: kfenn@msu.edu
The authors would like to thank the Division of Social Sciences at The University of Chicago for supporting this research, Uri Hasson and Susan Wagner Cook for valuable comments on the manuscript, and Gretchen Neidhardt, Nicole Donders, Matthew Tipping, Rachel Hilbert, and Michelle Chun for their assistance in conducting the study.

http://www.psypress.com/qjep
DOI:10.1080/17470218.2011.570353

Rensink, 2005). Perhaps more surprising, people often fail to detect a change in interlocutor during face-to-face interactions, when visual (face and body) information is available in addition to auditory (speech) information (cf. Levin, Simons, Angelone, & Chabris, 2002; Simons & Levin, 1998). Although much research has served to delineate the conditions under which observers detect, or fail to detect, changes in a visual scene, much less is known about the failure to detect changes in the auditory domain (for recent exceptions see Eramudugolla, Irvine, McAnally, Martin, & Mattingly, 2005; Gregg & Samuel, 2008; Pavani & Turatto, 2008; Vitevitch, 2003).

Speech presents an interesting avenue to investigate change detection. The speech signal is inherently complex and may vary on several different dimensions, all of which convey some information about the intended message. For example, the speech signal contains both a lexical–propositional dimension and a prosodic dimension that can convey attitudinal and emotional information as well as referential information (see Shintel, Nusbaum, & Okrent, 2006). In addition, speech contains source information, which includes information about talker identity, such as gender and age. The standard model of speech perception has always assumed that acoustic details of speech sounds are lost as a result of categorical perception (e.g., McQueen, Norris, & Cutler, 2006; Norris, McQueen, & Cutler, 2000; Studdert-Kennedy, Liberman, Harris, & Cooper, 1970) and that word recognition proceeds from the phonetic patterns (although see Grosjean & Gee, 1987, for a critique of this approach). Since the standard assumption that word recognition proceeds from phoneme or phonetic feature patterns, models of word recognition such as the cohort model (Marslen-Wilson & Welsh, 1978), the TRACE model (McClelland & Elman, 1986), and Shortlist (Norris, 1994) all represent the word recognition process as dependent solely on the phonetic structure of a spoken word without any consideration of the source of this information. This is similar to the way that the Logogen model represents words by constituent pattern properties (Morton, 1969). Furthermore, there is also evidence that semantic processing is independent of talker identity. Using a speech priming paradigm, Kouider and Dupoux (2005) found that changing the voice of the talker between the prime and the target did not affect repetition priming. Similarly, the reduction in neural response associated with repetition of a spoken word does not vary based on the voice of the talker (Orfanidou, Marslen-Wilson, & Davis, 2006). Given that speaker-specific information (such as the speaker's vocal characteristics) may be stripped away or normalized during the word recognition process, changes in source information might readily go unnoticed in a conversation.

In contrast, episodic accounts of speech perception proposed that listeners automatically encode indexical information, including the talker's vocal characteristics, along with the acoustic–phonetic properties of an utterance (e.g., Goldinger, 1998). These episodic traces are maintained in long-term memory as a natural part of speech perception. This suggests that selective attention to voice properties should not be necessary for talker change detection. Consistent with this, Kraljic and colleagues have shown that listeners are sensitive to different sources of acoustic variation in a talker's speech (vocal characteristics, idiolect, dialect) and must adjust accordingly (Kraljic, Brennan, & Samuel, 2008; Kraljic, Samuel, & Brennan, 2008). Thus, listeners should be sensitive to changes of talker within a conversation, since they may have to adjust perceptual processing in order to maintain phonological constancy.

The issue of encoding speaker-specific vocal information is especially relevant in contexts of voice-only communication. Because face information provides a key cue to personal identity, in the absence of such information listeners need to rely on voice information for inferring the identity of the speaker, a cue that is perhaps less distinctive (Read & Craik, 1995; Wilding & Cook, 2000; see Yarney, 2007, for review). Thus, a question arises as to whether listeners can readily detect voice changes and whether they spontaneously do so in the context of a natural conversation. Previous findings have shown that the same sentence can be interpreted differently when uttered by different speakers (Krauss & Fussell, 1991; Metzing & Brennan, 2003), suggesting that listeners' awareness (or lack thereof) of conversational changes could have significant conversational consequences.

One factor that is critical to noticing changes in speech or in talker vocal characteristics is attention. Several lines of research suggest that when attention is directed away from speech, the ability to detect information in that speech stream is reduced. Sinnett and colleagues (Sinnett, Costa, & Soto-Faraco, 2006) showed that word recognition accuracy was reduced when attention was directed away from the speech stream. Cherry (1953) found that when shadowing speech, directing attention to one stream of speech impairs detection of a change in language in the unattended stream (although changes in gender were rarely missed). For nonspeech auditory changes, Eramudugolla and colleagues (2005) demonstrated that in complex auditory scenes, when information arrived from many different sources and directions, listeners often missed changes in auditory objects unless attention was directed to those objects (similar to visual change detection, see O'Regan, Duebel, Clark, & Rensink, 2000; Rensink, O'Regan, & Clark, 1997).

Although changes that are not attended are rarely noticed, it remains unresolved whether changes in an *attended* speech stream will show similar change deafness. Gregg and Samuel (2008) argued that when changes in auditory objects are carried by acoustic differences in F0 (fundamental frequency) or harmonicity, which is claimed to be preattentively processed (Kat & Samuel, 1984; Samuel & Newport, 1979), changes are more likely to be detected. Unlike the aforementioned studies, which included auditory scenes with multiple objects (i.e., Eramudugolla et al., 2005; Gregg & Samuel, 2008; Sinnett et al., 2006), attending to a single stream of spoken words does direct attention to a specific auditory object. However, this auditory object is inherently complex and contains several dimensions of information. Thus, constraining attention to a single speech stream may not be enough to prevent change deafness. Vitevitch (2003) found change deafness to talker changes when participants shadowed spoken words in real time. Following a break in the experiment, 42% of listeners missed a change in talker that occurred between the first and second shadowing blocks. Vitevich argued that attending to the phonetic pattern of words resulted in disattention to voice quality. However, shadowing requires the subject to hear and rapidly repeat aloud the speech that is heard, resulting in significant cognitive demands that are measurably greater from simply listening to speech (e.g., Allport, Antonis, & Reynolds, 1972; Hyona, Tommola, & Alaja, 1995; Shallice, McLeod, & Lewis, 1985). Thus, it is possible that the change deafness observed in this study could have been due to the load of the shadowing task.

This raises the issue of processing talker information and talker changes in normal conversation. On the one hand, in carrying out a conversation, listeners face a theoretically complex task, albeit subjectively simple, requiring them to interpret syntactic, as well as lexical, information from the signal, to arrive at a pragmatically appropriate interpretation of the message, and to formulate a reply. Listeners may conserve processing resources and focus attention on spectrotemporal patterns of speech most relevant to linguistic interpretation and ignore talker information. In addition, listeners may assume that during conversation, the message changes but the conversational partner does not. Thus, they may not monitor their memory for the talker and may miss changes in talker. This suggests that talker change deafness in a normal conversation may even exceed the change deafness observed in Vitevitch's (2003) shadowing study, even though listeners are actively attending the voice in the conversation.

On the other hand, talker information may be more relevant in the context of a normal conversation. Given that the identity of the talker is irrelevant for shadowing, listeners may mobilize the resources available to them for task-relevant processes in response to the attentional load of shadowing (e.g., Allport et al., 1972; Conway, Cowan, & Bunting, 2001; Hyona, Tommola, & Alaja, 1995; Shallice et al., 1985). In contrast, listeners engaged in conversation must attend not only to the linguistic message, but presumably to other cues such as prosody, in order to interpret the message and respond appropriately. Evidence suggests that rather than representing information that is extrinsic to the message, talker information has a reliable effect on the interpretation of the message. For example, pedestrians' responses to a question asking for directions depended on the talker's regional accent (Kingsbury, 1968, reported in Krauss & Fussell, 1991), and referring expressions were interpreted differently when produced by different talkers (Metzing & Brennan, 2003). Furthermore, there is some evidence that during comprehension, information about the speaker's identity is integrated with information about the message. Sentences whose semantic content was inconsistent with voice-conveyed talker characteristics (such as gender or age) elicited an N400 response, similar to that elicited by semantic anomalies (Lattner & Friederici, 2003; Van Berkum, van den Brink, Tesink, Kos, & Hagoort, 2008), and were associated with enhanced activity in the inferior frontal gyrus (Tesink et al., 2009). These findings suggest that listeners may attend to talker information in normal conversation. Although conversation structure does impose some constraints on the timing of the response, it may be flexible enough to allow listeners to delay responses, thereby leaving more attention to devote to the talker's voice. In this case, we would not expect significant change deafness, since listeners are specifically attending to acoustic information that conveys talker identity.

We investigated talker change detection in a phone conversation. Participants were called on the telephone to participate in an olfactory memory survey. Halfway through the conversation, the talker changed, and a new experimenter completed the survey. Voice memory was later tested.

GENERAL METHOD

Participants

For all experiments, participants were students or employees at the University of Chicago. All were native speakers of American English and reported no history of hearing or speech disorders. All were paid $10, except where noted.

Materials

The same survey, consisting of 12 questions regarding odour preferences and associations, was used for all participants (Appendix). All voice test stimuli were utterances produced by the experimenters and were digitized and sampled at 44.1 kHz with 16-bit resolution. Sounds were played directly from the computer soundcard through the phone using a Voice Path adapter by JK Audio.

Procedure

The order of survey questions was fixed but the order of experimenters was counterbalanced across participants within each experiment. Participants were recruited by e-mail and were called on the phone to participate. Verbal consent was obtained to tape the conversation and for participation in the experiment. All participants were debriefed regarding the actual purpose of the study.

EXPERIMENT 1

The purpose of the first experiment was to test whether or not listeners would detect changes in talker during a normal phone conversation. This experiment also tested memory for the two voices that were heard in the conversation. Several theories have been proposed to account for the lack of change detection in vision (see Simons, 2000), and testing the two voices with each other may help us to better understand a lack of change detection in speech. For example, if memory for the first voice is better than memory for the second voice, this might suggest that the second voice was simply not encoded. Alternatively, if memory for the second voice is better than memory for the first, this might suggest that the second voice may have overwritten (or interfered with) memory for the first voice. In visual change detection, Levin and colleagues (2002) found that participants who detected a change in interlocutor could reliably identify the original person in a subsequent recognition task, but those who failed to detect the change performed at chance, suggesting that the individuals who did not detect a change may have failed to encode the first person, or that the image of the second person may have overwritten memory for the first.

However, later change blindness studies have shown that even when participants miss a change, they are often able to remember prechange information, suggesting that both prechange and post-change visual scenes may be encoded and remembered, but may not be compared to each other. Simons and colleagues (2002) found that participants who did not notice the removal of an object from a scene (e.g., a red ball) still retained some memory of the object, indicating that these details were encoded and stored in memory. A related study found that participants who did miss a person change in a video experiment *were* able to reliably identify the first person in a line-up (Angelone, Levin, & Simons, 2003), and similar memory for prechange information has been found in a two-dimensional change detection paradigm (Mitroff, Simons, & Levin, 2004), suggesting that this information was both encoded and remembered.

Method

Participants

Sixteen participants completed this experiment (mean age = 19.3 years; 11 female).

Materials

Individual voice samples were recorded of the two experimenters asking the same question ("What smell do you think about when you're angry?") for the voice memory test.

Procedure

The first experimenter called the participant, explained the study, and obtained consent. Directly after she had read the third question in the survey (Appendix), the other experimenter took the phone and exclaimed: "Oh wait a minute, I read the wrong question. The questions must be answered in a specific order." This manipulation was designed to minimize the interval between the two voices and to perform the change when the participant was focused on the conversation and presumably on the voice of the talker. The average duration between voices was about one second. Two female experimenters conducted testing. The mean F0s for these experimenters, averaged over several randomly selected sentences, were 218 Hz and 200 Hz. Because we wanted to compare memory for the two voices, we aimed to match the time spent talking to each experimenter. The average duration of the first half of the experiment (when participants spoke with the first experimenter) was 2.32 ± 0.36 minutes, and the average duration of the second half was 2.45 ± 0.8 minutes, a difference that was not significant $t(15) = 0.75$, $p > .46$. Thus, listeners had roughly equal experience with both voices.

Voice recognition test. At the conclusion of the survey, the experimenter told participants that her "supervisor" needed to ask some follow-up questions. To avoid confusion of voice memory, the third experimenter was a heavily accented native Hebrew speaker and did not sound similar to either experimenter. The supervisor explained that the experiment was testing voice memory in addition to odour memory. The test used a two-alternative forced-choice task (2AFC). Participants heard two voices asking the same question, were presented with a 500-ms interval between them, and were asked to choose the voice of the person with whom they had just spoken. Voice order at test was counterbalanced with voice order during the survey. A stronger memory trace for the first or the second voice should result in greater probability of choosing that voice in the forced-choice memory task. The test was conducted prior to debriefing so participants did not know that they had spoken to two people at this point.

After the test, the supervisor asked increasingly specific questions to probe whether participants had detected the switch (cf. Levin et al., 2002; Simons & Levin, 1998). The four questions were: (a) "Did anything unusual or strange happen during the interview?"; (b) "Did anything change during the interview?"; (c) "Was there anything unusual about the experimenter's voice during the interview?"; (d) "Did the experimenter's voice change at all during the interview?". The supervisor then debriefed the participants.

Results

Only 1 individual (6% of participants) noticed the changed voice. This participant remarked that the voice "started off louder and different sounding" when asked whether there was anything unusual about the experimenter's voice. One other participant remarked that the experimenter sounded louder after the switch point, but did not notice anything different about the voice quality. Although the average interval between voices was greater than the interval used in many change blindness studies, the auditory memory trace is believed to last for several seconds, suggesting that memory for the first voice should be maintained. Furthermore, the interval between the voices was blank, so change deafness could not be explained by a filled interval erasing the memory trace.

In the voice recognition test, we were interested in whether the participants who did not notice a change (i.e., 15 out of 16 participants) would have a greater tendency to report that either the first or the second voice was the person with whom they spoke. In this test, 9 participants (60%) chose the voice of the first experimenter, and 6 (40%) chose the second experimenter, a difference that is not significant when tested with a two-tailed sign test ($p = .6$). There were no effects of experimenter voice; across different orders, 7 participants chose the voice of one experimenter, and 8 chose the voice of the other experimenter, suggesting that results did not reflect a preference for a specific experimenter's voice.

These results suggest that when engaged in a naturalistic conversation, listeners often fail to detect a change in conversational partner, even though they must actively attend the message and therefore the speech in order to respond appropriately. Listeners may not continuously monitor the vocal properties of their interlocutor. Change detection rates in this study were quite low in comparison to earlier reports of change detection in speech. Only about 6% of the participants noticed the changed voice in this experiment. In contrast, Vitevitch (2003) showed that when shadowing speech, 58% of listeners noticed a change in voice. This is rather surprising as it suggests that when listeners are actively engaged in a conversation with another person, they are actually less likely to notice a change in voice than when they are shadowing speech.

In this study, listeners seem to remember both the first and the second voices at approximately the same rate. Although there appears to be a slight trend for participants to choose the first voice, there is no clear pattern for which voice participants are more likely to choose. There are several possible explanations for this result. This could suggest either that listeners are initially encoding fine details of both voices but are retaining only *some* memory for these auditory properties, or that listeners are encoding only *some* auditory properties of both voices and are retaining memory for these incomplete representations. Voice memory may not be an all-or-none phenomenon; listeners may retain some sort of general memory for a voice or may remember only certain aspects of the voice and not retain a fully specified, veridical memory. In other words, participants may retain only an auditory "gist" memory of voices. For example, they may encode the voice as being female and in the same age range as themselves (i.e., young adult vs. child or older adult). This sort of general memory might not allow them to detect the change or reliably decide which voice was the voice most recently heard as they both sound familiar in some way (i.e., both female and young adult).

However, before accepting the conclusion that these findings reflect some memory for both voices, we need to reject one alternative interpretation of the results: It is possible that the lack of clear preference for either voice does not reflect memory for both, but results from chance performance. Participants may not remember the voices at all and may simply be choosing a voice at random, because they are required to make a choice. Given that voice memory is not particularly strong—listeners are generally quite poor in "voice line-ups" (cf. Read & Craik, 1995; Wilding & Cook, 2000; see Yarney, 2007, for review)—it is possible that participants simply did not remember the voices they heard and thus showed no preference for either voice. If this is the case, we would expect to observe the same pattern of results even if one of the voices presented in the recognition test had not been presented during the survey at all. This possibility is examined in Experiment 2.

EXPERIMENT 2

Experiment 2 was designed to test memory for a voice that was encountered in the conversation compared to a voice that had not been encountered during the conversation. If participants retain some memory of both presented voices, but this memory is not sufficient for change detection, they should be able to reliably discriminate between a heard and an unheard voice. On the other hand, if participants do not remember either voice (and thus the results of Experiment 1 reflect chance performance), they should not be able to discriminate between a heard and an unheard voice. We used the same methodology as that in the previous experiment; however, during the test at the end of the experiment we presented only one of the two experimenters' voices contrasted with a previously unheard talker.

Method

Participants

The 24 participants had a mean age of 21.7 years, and 19 were female.

Materials

Individual voice samples were recorded of the experimenters asking the same question as that in Experiment 1.

Procedure

Three experimenters conducted testing. One of the experimenters, ($F0 = 218$ Hz) also conducted Experiment 1; the other two experimenters were new. The mean F0 for the two new experimenters was 202 Hz and 239 Hz, respectively. The average duration of the first half of the experiment was 2.3 $\pm$ 0.33 minutes, and the average duration of the second half was 2.6 $\pm$ 0.88 minutes, a marginally significant difference, $t(23) = 1.88$, $p = .07$.

The procedure used in Experiment 1 was also used in this experiment with the exception of the voices used in the test. In the 2AFC test at the end of the experiment, one test voice was the voice of the one of the experimenters who conducted the survey (either the first or the second experimenter), and the other voice was the voice of a third experimenter, with whom that particular individual had never spoken. Voices were presented during the test with a 250-ms interval between them.

Each of the three experimenters conducted testing for two thirds of the participants, and experimenter voices and voice order were counterbalanced across participants. Test voice (first or second experimenter), voice order at test (experimenter voice first or unheard voice first), and voice identity (Experimenter 1, 2, or 3) were also counterbalanced. The test and the debriefing were conducted by another experimenter (the same "supervisor" as that in Experiment 1). Finally, the consent form was prerecorded and played for the participants through the phone, instead of being read aloud to participants. The consent form was always played in the voice of the experimenter who was currently speaking to the participant.

Results

Only one person (4%) of the 24 tested noticed the voice change. In response to the second probe question, this person correctly remarked that the voice changed when the experimenter asked the wrong question.

In the voice test, 17 (74%) of the participants who did not notice the change correctly identified the experimenter, significantly above chance (two-tailed sign test, $p < .05$). Only 6 (26%) believed that the unheard voice was the voice of the experimenter

The proportion of correct voice recognition did not differ significantly between the group that was tested on the first experimenter's voice (10 out of 12 correct) and the group that was tested with the second experimenter's voice (7 out of 11 correct), $\chi^2(1) = 1.15$, $p > .28$. Two of the people who incorrectly chose the nonpresented voice were tested with the voice that occurred first in the conversation, and 4 were tested with the second voice.

These results suggest that although participants are not more likely to recognize either voice, they encoded and retained some information about each talker—enough to distinguish between heard and unheard talkers. It is important to note that because the three voices were completely counterbalanced both during the survey and at test, and because change detection performance was uniformly low across voices, these results could not be due to the unheard voice simply being more dissimilar to the other voices. However, it is also possible that not all participants are encoding and remembering both voices, but that some participants remembered the first voice, and some participants remembered the second voice, but none of the participants actually remembered both voices. While the present findings cannot rule out this possibility, they do argue against a systematic retention of either the initial or the latter representation that is consistent across participants. Finally, these results are also consistent with the possibility, suggested by change blindness research, that information about both voices is encoded and retained, but they are not spontaneously compared to each other (for supporting evidence for this explanation in the visual domain see Angelone et al., 2003; Mitroff et al., 2004; Simons et al., 2002).

EXPERIMENT 3

The results of the previous experiments suggest that people do not automatically detect voice changes in conversation, even when they retain some memory for the voices. If the voices are sufficiently similar, detecting a change while carrying out a conversation may simply be too demanding given the cognitive demands of carrying out a phone conversation. On the other hand, change deafness may not reflect inherent difficulty but rather listeners' expectations and strategies in a specific context; voice changes may not have been detected in the earlier experiments because participants did not expect a talker change in the middle of a conversation. When talking on the phone, it is not uncommon to have interruptions during the call. When the conversation resumes after an interruption, we assume the person on the phone will be the same person. If participants assume that they are talking to one person, they may ignore voice properties that are inconsistent with this belief or interpret voice characteristics in a way that confirms their expectations. If this is the case, alerting participants to the possibility of change may facilitate detection. Indeed, a related study on complex auditory scenes found that directing listeners' attention to the spatial location of a potential change facilitated auditory change detection (Eramudugolla et al., 2005), suggesting that focusing listeners' attention on acoustic properties of the voice may increase change detection in our task. The fourth experiment tested whether warning participants that a change may occur, thus directing their attention to voice characteristics, would increase change detection rates.

Method

Participants

Twelve participants completed this experiment (mean age = 19.3 years; 7 female).

Procedure

The experimenters who conducted Experiment 1 also conducted this experiment, and the procedure used in Experiments 1 and 2 was also used, with modifications. First, before the survey began, participants were told that either they would speak to one person throughout the survey or they would speak to two different people. Therefore, they were told that there might be a change in talker.

They were further told that a change such as this would occur for half of the participants (in fact a change always occurred). Participants were instructed to interrupt the conversation if they detected a change. To ensure participant motivation, a $2 incentive was offered to anyone who either correctly detected a change in talker or correctly detected that no change occurred (all participants were paid the additional $2, regardless of detection performance). At the conclusion of the survey, a third experimenter (the "supervisor" from Experiments 1 and 2) came onto the phone and asked each participant how many people conducted the survey.

Results

In contrast to the low rates of change detection found in the first three experiments, when told that a talker change might occur during the conversation, participants reliably reported the change. Nine of the 12 participants (75%) reported talking to two different people. Six of these people interrupted the conversation at the switch point to report the changed voice; 4 participants reported the changed voice immediately when the second experimenter came onto the phone, and the other 2 reported it directly after the second experimenter read her first question. It is important to note that given that participants could have interrupted the survey at any point, the fact that they identified the correct switch point shows that their response reflected improved change *detection* rather than merely increased likelihood of change *reporting*. The remaining 3 participants did not interrupt the conversation, but during questioning reported that they had spoken to two different people. This change detection rate is significantly higher than that in Experiment 1, $\chi^2(1) = 14.1$, $p < .001$, and Experiment 2, $\chi^2(1) = 20$, $p < .0001$. Furthermore, even if we take a conservative stance and calculate these statistics using only the 6 participants who spontaneously reported the changed voice at the moment of the change, we still see significantly greater change detection in this experiment than in the other experiments: Experiment 1: $\chi^2(1) = 10.4$, $p = .001$; Experiment 2: $\chi^2(1) = 15.3$, $p < .0001$. The significant increase in detection relative to Experiment 1 is especially informative, given that both involved the same voices. Thus, the increased detection rate cannot be attributed to methodological differences between the experiments or differences in the voices.

These results suggest that participants are able to carry out a conversation and successfully monitor for a talker change. Clearly the conversation did not impose a cognitive load that would prevent talker change detection. Rather, this supports the idea that listeners do not generally expect a talker change in a conversation and may not typically listen to the voice quality of a talker at a level of resolution needed to recognize a talker change. When this general expectation is suspended, change detection increases. This raises a question regarding how expectations about conversations might influence the way that listeners interpret voice information. We explored this question in Experiment 4.

EXPERIMENT 4

In Experiment 3, listeners were made aware of the possibility of change, while keeping the change itself the same as that in the previous experiments. The results suggest that expectations regarding talker changes may have governed the way listeners attended to the speech and thus affected their ability to detect a change. This finding is suggested by previous research by Magnuson and Nusbaum (2007). Talker variability has been found to slow recognition (Mullennix & Pisoni, 1990) and increase working memory load (Nusbaum & Morin, 1992). However, the effect of talker variability is not fixed but rather is dependent on listeners' expectations. In one study (Magnuson & Nusbaum, 2007), when listeners expected variation in pitch to reflect a talker difference, there was evidence of talker normalization reflected in increased processing costs. When expecting a change in talker, listeners' attention could be directed towards a reliable acoustic difference (a small change in F0 with no change in vocal tract) between voices, and this served to trigger normalization processing. However, when the same acoustic difference was expected to reflect a single talker who used the pitch difference to highlight words, talker normalization did not occur (accuracy was not impaired). Magnuson and Nusbaum maintained that these results argue for an active control of processing in speech perception, in which the interpretation of the speech signal is constrained by both bottom-up and top-down factors. To investigate whether expectations can play a similar role in talker change detection, we reversed listeners' expectations regarding the presence of a change and manipulated the actual occurrence of a change. All participants were told that there would be a talker change during the experiment; for half the participants the talker did change, and for half the talker did not change.

Method

Participants

The 20 participants (mean age = 20.8 years; 11 female) were randomly assigned to either the same-talker or the changed-talker condition.

Procedure

Two female experimenters conducted this experiment. One experimenter also conducted Experiments 1–3 (mean F0 of 218 Hz), and one was new (215 Hz). Experimenter (in the same-talker condition) and order of experimenters (in the changed-talker condition) were counterbalanced across participants.

The procedure was identical to that of Experiments 1 and 2 except for the following: Prior to starting the experiment, the experimenter told participants that she would ask six questions, and then a different experimenter would come on the phone. After the first six questions, the experimenter placed the participant on hold and reiterated that the "other" experimenter would finish the survey. After a pause that averaged 6 seconds, either a new experimenter (changed talker) or the same experimenter (same talker) returned and finished the survey. This change procedure and delay interval were used in several (unreported) experiments that showed average change detection rates of approximately 15–20%.

At the completion of the survey, all participants were asked a series of four questions: (a) "Did anything unusual or strange happen during the interview?"; (b) "Was there anything unusual about the first experimenter's voice? (If yes, please explain)"; (c) "Was there anything unusual about my [the current experimenter's] voice? (If yes, please explain)"; (d) "Is my voice different from the first experimenter's?".

Results

All participants in the changed-talker group (n = 10) reported that the two voices were different, but 1 participant remarked that he thought that they were different, but was not sure because he wasn't paying close attention. Although the participants in the same-talker condition (n = 10) expected a change in experimenter, none stopped the conversation when the same experimenter returned, and all finished the survey without comment. Furthermore, during the survey, 3 participants distinguished between the "first" and "second" experimenters, referring back to the "first" or the "other" experimenter.

During the course of questioning, only 2 participants (20%) in the same-talker group answered the probe questions in a manner that suggested that they did not know whether the voices were different. When asked whether the current experimenter's voice was different from the first experimenter's, 1 participant reported that he "honestly didn't know". The other participant reported that she "assumed" that the voices were different but that she wasn't paying attention and did not really know. This response is similar to that of the aforementioned participant in the changed-talker group.

These results further support the idea that the way that listeners attend to and encode information seems to be influenced by their expectations. Indeed, Rubin (1992) has found that listeners' expectations of talker identity significantly affect the way in which speech is perceived and evaluated. Participants listened to prose passages recorded by a native American English speaker and were given a picture supposedly representing the talker. When the picture depicted an Asian woman, participants rated the speech as more accented than when the picture depicted a Caucasian woman, although both groups listened to the same speech (Rubin, 1992, Experiment 1). This suggests that listeners may be attending to voice properties differently, based on their expectations of the speech. Similarly, in the present experiment, people may interpret acoustic–phonetic differences according to their expectations. Small differences may be interpreted as random within-talker differences when listeners do not consider the possibility of a talker change, but as indicating a different talker when such a change is expected.

EXPERIMENT 5

The previous experiments provide strong evidence that people do not often spontaneously notice a change in talker, despite their ability to detect such a change. In this experiment, we wanted to investigate what sort of changes people spontaneously notice. We predicted that a very salient change that requires acoustic adaptation would be more readily detected. We tested the extent to which participants would notice a change in talker if the two talkers differed in gender.

Method

Participants

The 12 participants had a mean age of 19.3 years (4 female).

Procedure

Two experimenters conducted this experiment. One experimenter was female and also conducted testing in Experiment 3; the other experimenter was male and had not conducted testing previously. The same basic procedure was used as that in the previous experiments except that the experimenters switched while the participant answered the fourth question in the survey. This was the same point in the survey at which the switch occurred in Experiments 1–3, but there was not an immediate change in voice. This modification was introduced because we assumed that we would find very high rates of change detection and wanted to make the task slightly more difficult.

Results

Not surprisingly, 11 out of 12 participants (92%) noticed the change in gender. This is significantly more than even the highest rate of spontaneous detection of same-gender voice change (in Experiment 1), $\chi^2(1) = 16.9$, $p < .0001$. None of the participants interrupted the conversation to report the changed voice, but seemed to accept the voice change as part of the experiment. When asked the probe questions, the 11 participants who noticed the change all remarked that nothing was different or changed but commented that the only change they observed was that there was a female talker and male talker during the conversation. The one participant who did not notice the change did not remark on any sort of change in voice and seemed quite surprised that he had been speaking to two people of different genders.

These results are similar to Cherry's (1953) finding that changes in gender are reliably noticed even if they occur in an unattended channel. In this experiment, we sought to extend the findings of our previous studies to show that gender changes are salient when engaged in natural conversation. Although this is not a particularly surprising finding, it does show that extreme voice changes are easily detected. Furthermore, if participants indeed encode and retain a gist memory for the voice (i.e., female, young adult) then a change in gender (to male) would deviate sufficiently from the gist representation to enable change detection. It is possible that a change in gender would not require that the original voice be compared to the changed voice for detection as gender is, in some ways, the most central attribute of a voice. For analogy, changes that occur in more central areas of a visual scene tend to be noticed more often than changes in the more peripheral areas (Rensink et al., 1997).

GENERAL DISCUSSION

During phone conversations, listeners are not sensitive to unexpected changes in their conversational partner: When a new person unexpectedly continued a conversation, participants rarely noticed, although they had the perceptual sensitivity to detect this change. This finding has important implications for theories on speech perception. Although some recent models suggest that talker vocal characteristics are automatically encoded with the acoustic–phonetic properties of an utterance (cf. Goldinger, 1998), our results suggest that listeners may not automatically encode both the conversational message and voice characteristics when listening to speech. Instead, when engaged in normal conversation, listeners may focus attention on understanding the message and may perceive speech based solely on the phonetic pattern structure of a spoken word without processing the source information or voice that conveyed the information (cf. Marslen-Wilson & Welsh, 1978; McClelland & Elman, 1986; Norris, 1994). If language use evolved in service of face-to-face conversation (cf. Pickering & Garrod, 2004; Rizzolatti & Arbib, 1998), and conversational partners typically do not appear and disappear abruptly during the conversation, then there is no reason for listeners to continuously monitor the identity of their interlocutor. Moreover, our daily experience with conversation indicates that typically when a new talker enters a conversation there are spatial cues and conversational pragmatic cues that signal the introduction of the talker. Thus, there is no reason for language processing to develop an alarm mechanism that would continuously monitor the talker's identity and automatically signal a talker change. Given the assumption of interlocutor stability, listeners are free to focus attention on the linguistic message.

These results further suggest that expectations may affect the way listeners attend to speech and in particular to voices, in conversation, and that listeners do not habitually attend to all perceptually available information. Changes in voice are not detected automatically and spontaneously, even though participants were able detect the change when the possibility of a talker change was signalled prior to a conversation. Small acoustic differences may be attended to and interpreted as talker differences, if a change in talker is expected (cf. Magnuson & Nusbaum, 2007). Listeners may, according to a priori expectations, monitor the sound of a talker's voice at different levels of analysis. When expecting that a change in talker might occur, listeners may attend to voice properties to monitor for a change and may become more sensitive to differences between voices or even more likely to interpret within-talker acoustic variability as evidence for a talker change. However, when there is no reason to expect a talker change, voice information may be perceived or monitored at a relatively coarse level of analysis; talker information may not be encoded or may be interpreted in a manner that confirms the listener's expectations that their interlocutor is the same person. However, when talkers differ in vocal tract sufficiently, such as when talkers differ in gender, these bottom-up acoustic differences may grab attention even in the absence of a top-down expectation of change. Indeed, this type of dramatic change was rarely missed, even without expectation of a change. Therefore, detecting a talker change may require a relatively large acoustic difference between talkers, beyond the range of a single talker's vocal variability.

This is supported by evidence that some properties of the two voices were encoded and remembered even when a change was not detected. Participants could reliably identify each experimenter's voice and distinguish it from a new talker's voice, despite a lack of change detection. However, participants were not able to determine which voice was the most recent (postchange) voice. These results suggest that participants encode some information about both voices and retain some memory for both of them, or that some participants retain information about the first voice, and some retain information about the second voice. However, they also suggest that the retained memory may be weak or incomplete, with respect to specific acoustic characteristics of the voice or contextual details. If memory for the voices is a somewhat vague, or a gist-based representation in which listeners maintain a general idea of the voice as being of a certain gender and certain age range, this may explain the low change detection rates. Listeners may not detect a change even if they were to compare the two voices if both voices fall into the same broad category.

Taken together, these findings suggest that talker change deafness is not a result of deterministic fixed processing limitations that prevent attention to talker indexical information while engaging in conversation. Instead, deafness to talker change appears to result from the way that our expectations about the conversation shape the way we direct attention to speech. Such an active control system (Magnuson & Nusbaum, 2007) would allow for flexible processing, adapted to the specific context. When expecting a speaker change, listeners can shift their attention to fine-grained voice information. But under normal circumstances, given the typical expectation that a conversational partner will be stable, interlocutors can safely focus attention on the spectrotemporal properties of spoken words, which serves the purpose of carrying on a conversation. By focusing attention on what is being said rather than on who is saying it, once we know who is talking, we can marshal cognitive resources for the other aspects of a conversation.

Original manuscript received 16 April 2010
Accepted revision received 17 February 2011
First published online 20 May 2011

REFERENCES

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

Angelone, B. L., Levin, D. T., & Simons, D. J. (2003). The relationship between change detection and recognition of centrally attended objects in motion pictures. *Perception*, *32,* 947–962.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and two ears. *Journal of the Acoustical Society of America*, *25,* 975–979.

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

Eramudugolla, R., Irvine, D. R. F., McAnally, K. I., Martin, R. L., & Mattingly, J. B. (2005). Directed attention eliminates "change deafness" in complex auditory scenes. *Current Biology*, *15,* 1108–1113.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105,* 251–279.

Gregg, M. K., & Samuel, A. G. (2008). Change deafness and the organizational properties of sounds. *Journal of Experimental Psychology: Human Perception and Performance*, *34,* 974–991.

Grimes, J. (1996). On the failure to detect changes in scenes across saccades. In K. Akins (Ed.), *Perception: Vol. 2. Vancouver studies in cognitive science* (pp. 89–110). New York, NY: Oxford University Press.

Grosjean, F., & Gee, J. (1987). Prosodic structure and spoken word recognition. *Cognition*, *25,* 135–155.

Hollingworth, A. (2004). Constructing visual representations of natural scenes: The roles of short- and long-term visual memory. *Journal of Experimental Psychology: Human Perception & Performance*, *30*(3), 519–537.

Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception & Performance*, *28*(1), 113–136.

Hyona, J., Tommola, J., & Alaja, A.-M. (1995). Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *Quarterly Journal of Experimental Psychology*, *48A*(3), 598–612.

Kat, D., & Samuel, A. G. (1984). More adaptation of speech by nonspeech. *Journal of Experimental Psychology: Human Perception and Performance*, *10,* 512–525.

Kingsbury, D. (1968). *Manipulating the amount of information obtained from a person giving directions.*

Unpublished honours thesis, Department of Social Relations, Harvard University, Cambridge, MA.

Kouider, S., & Dupoux, E. (2005). Subliminal speech priming. *Psychological Science*, *16*(8), 617–625.

Kraljic, T., Brennan, S. E., & Samuel, A. G. (2008). Accommodating variation: Dialects, idiolects, and speech processing. *Cognition*, *107*(1), 54–81.

Kraljic, T., Samuel, A. G., & Brennan, S. E. (2008). First impressions and last resorts: How listeners adjust to speaker variability. *Psychological Science*, *19*(4), 332–338.

Krauss, R. M., & Fussell, S. R. (1991). Perspective-taking in communication representation of others' knowledge in reference. *Social Cognition*, *9*(1), 2–24.

Lattner, S., & Friederici, A. D. (2003). Talker's voice and gender stereotype in human auditory sentence processing—Evidence from event-related brain potentials. *Neuroscience Letters*, *339*(3), 191–194.

Levin, D. T., Simons, D. J., Angelone, B. L., & Chabris, C. F. (2002). Memory for centrally attended changing objects in an incidental real-world change detection paradigm. *British Journal of Psychology*, *93,* 289–302.

Magnuson, J. S., & Nusbaum, H. C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology: Human Perception and Performance*, *33,* 391–409.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, *10,* 29–63.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18,* 1–86.

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

Metzing, C., & Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory & Language*, *49,* 201–213.

Mitroff, S. R., & Simons, D. J. (2002). Changes are not localized before they are explicitly detected. *Visual Cognition*, *9*(8), 937–968.

Mitroff, S. R., Simons, D. J., & Levin, D. T. (2004). Nothing compares 2 views: Change blindness can occur despite preserved access to the changed information. *Perception & Psychophysics*, *66*(8), 1268–1281.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, *76,* 165–178.

Mullennix, J. W., & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics*, *47,* 379–390.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52,* 189–234.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, *23,* 299–370.

Nusbaum, H. C., & Morin, T. M. (1992). Paying attention to differences among talkers. In Y. Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka (Eds.), *Speech perception, production, and linguistic structure*. Tokyo, Japan: OHM Publishing Company.

O'Regan, J. K., Duebel, H., Clark, J. J., & Rensink, R. A. (2000). Picture changes during blinks: Looking without seeing and seeing without looking. *Visual Cognition*, *7*(1), 191–211.

O'Regan, J. K., Rensink, R. A., & Clark, J. J. (1999). Change-blindness as a result of "mudsplashes". *Nature*, *398,* 34.

Orfanidou, E., Marslen-Wilson, W. D., & Davis, M. H. (2006). Neural response suppression predicts repetition priming of spoken words and pseudowords. *Journal of Cognitive Neuroscience*, *18,* 1237–1252.

Pavani, F., & Turatto, M. (2008). Change perception in complex auditory scenes. *Perception and Psychophysics*, *70*(4), 619–629.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27,* 169–226.

Read, D., & Craik, F. I. M. (1995). Earwitness identification: Some influences on voice recognition. *Journal of Experimental Psychology: Applied*, *1*(1), 6–18.

Rensink, R. A. (2002). Change detection. *Annual Review of Psychology*, *53*(1), 245–277.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, *8*(5), 368–373.

Rizzolatti, G., & Abib, M. A. (1998). Language within our grasp. *Trends in Neurosciences*, *21,* 188–194.

Rubin, D. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education*, *33*(4), 511–530.

Samuel, A. G., & Newport, E. L. (1979). Adaptation of speech by non-speech: Evidence for complex acoustic cue detectors. *Journal of Experimental Psychology: Human Perception and Performance*, 5, 563–578.

Scholl, B. J. (2000). Attenuated change blindness for exogenously attended items in a flicker paradigm. *Visual Cognition*, *7*(1), 377–396.

Shallice, T., McLeod, P., & Lewis, K. (1985). Isolating cognitive modules with the dual-task paradigm: Are speech perception and production separate processes? *Quarterly Journal of Experimental Psychology*, *37A,* 507–532.

Shintel, H., Nusbaum, H. C., & Okrent, A. (2006). Analog acoustic expression in speech communication. *Journal of Memory and Language*, *55*(2), 167–177.

Simons, D. J. (1996). In sight, out of mind: When object representations fail. *Psychological Science*, *7*(5), 301–305.

Simons, D. J. (2000). Current approaches to change blindness. *Visual Cognition*, *7,* 1–15.

Simons, D. J., Chabris, C. F., Schnur, T., & Levin, D. T. (2002). Evidence for preserved representations in change blindness. *Consciousness and Cognition*, *11,* 78–97.

Simons, D. J., Franconeri, S. L., & Reimer, R. L. (2000). Change blindness in the absence of a visual disruption. *Perception*, *29,* 1143–1154.

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences*, *1,* 261–267.

Simons, D. J., & Levin, D. T. (1998). Failure to detect changes to people during a real-world interaction. *Psychonomic Bulletin and Review*, *5*(4), 644–649.

Simons, D. J., & Rensink, R. A. (2005). Change blindness: Past, present, and future. *Trends in Cognitive Sciences*, *9*(1), 16–20.

Sinnet, S., Costa, A., & Soto-Faraco, S. (2006). Manipulating inattentional blindness within and across sensory modalities. *The Quarterly Journal of Experimental Psychology*, *59,* 1425–1442.

Studdert-Kennedy, M., Liberman, A. M., Harris, K. S., & Cooper, F. S. (1970). Motor theory of speech perception: A reply to Lane's critical review. *Psychological Review*, *77,* 234–249.

Tesink, C. M., Petersson, K. M., van Berkum, J. J., van den Brink, D., Buitelaar, J. K., & Hagoort, P. (2009). Unification of speaker and meaning in language comprehension: An fMRI study. *Journal of Cognitive Neuroscience*, *21*(11), 2085–2099.

Van Berkum, J. J. A., van den Brink, D., Tesink, C.M. J. Y., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, *20,* 580–591.

Vitevitch, M. S. (2003). Change deafness: The inability to detect changes between two voices. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(2), 333–342.

Wilding, J., & Cook, S. (2000). Sex differences and individual consistency in voice identification. *Perceptual and Motor Skills*, *91*(2), 535–538.

Yarney, A. D. (2007). The psychology of speaker identification and earwitness memory. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *The handbook of eyewitness psychology: Vol. II. Memory for people* (pp. 101–136). Mahwah, NJ: Lawrence Erlbaum Associates.

APPENDIX

Questions used in odour memory survey

In Experiments 1–3, the first experimenter asked Questions 1, 2, and 4. Immediately after the first experimenter had finished reading Question 4, the second experimenter came onto the phone and told the participant that she had read the incorrect question and asked Question 3. This experimenter then finished the survey. In Experiments 4, one experimenter asked the first six questions and then placed the participants on hold. A second experimenter then came onto the phone, resumed the survey with Question 7, and completed the survey. In Experiment 5, the first experimenter asked the first four questions. The experimenters switched when the participant was answering the fourth question.

1. What did your grandmother's house smell like?
2. What odour reminds you of your childhood?
3. What odour do you think of when you think about your elementary school?
4. What odour do you think of when you think about your high school?
5. What memories do you have that are closely associated with specific odours?
6. Why do you believe that odour is a salient characteristic of these memories?
7. What odours do you associate with specific seasons?
8. What odour reminds you of the holidays?
9. What is your favourite scent?
10. What is your least favourite scent?
11. What is your favourite fruit scent?
12. Are there any odours that you associate with specific emotions?