

Slow change deafness

John G. Neuhoff · Joseph Wayand ·
Mamoudou C. Ndiaye · Ann B. Berkow ·
Breanna R. Bertacchi · Catherine A. Benton

Published online: 19 March 2015
© The Psychonomic Society, Inc. 2015

Abstract In four experiments, we demonstrated a new phenomenon called “slow-change deafness.” In Experiment 1 we presented listeners with continuous speech that changed three semitones in pitch over time, and we found that nearly 50 % failed to notice the change. Experiments 2 and 3 replicated the finding, demonstrated that the changes in the stimuli were well above threshold, and showed that when listeners were alerted to the possibility of a change, detection rates improved dramatically. Experiment 4 showed that increasing the magnitude of the change that occurred in the stimulus decreased the rate of change deafness. Our results are consistent with previous work that had shown that cueing listeners to potential auditory changes can significantly reduce change deafness. These findings support an account of change deafness that is dependent on both the magnitude of a stimulus change and listener expectations.

Keywords Audition · Divided attention · Inattention

Seemingly obvious changes in our environment can sometimes go unnoticed. *Change blindness* and *change deafness* research has shown that observers can fail to notice when large objects change in a visual scene (Rensink, O’Regan, & Clark, 1997), when sounding objects enter and leave an auditory scene (Gregg & Samuel, 2008), when a male voice changes to female (Fenn et al., 2011), and even when an entire person is replaced with another during a real-life conversation (Simons & Levin, 1998). These examples illustrate a line of

change detection research that has provided insight into the nature of the attention, encoding, and cognitive representation that occur when we encounter auditory and visual objects (Rensink, 2002; Simons & Rensink, 2005; Vitevitch, 2003).

Though the principles and mechanisms that underlie the failure to detect suprathreshold change are still topics of considerable debate, the phenomena that have been discovered are remarkable. Some early work in vision examined change blindness when images were changed during a saccade. For example, Grimes (1996) showed that viewers can fail to notice when two people in an image exchange heads. In later work using the “flicker paradigm,” Rensink et al. (1997) showed that even changes that did not occur during a saccade often went unnoticed if the images were separated by a very brief masking stimulus. Simons and Levin (1998) even showed that participants often failed to notice when an entire person was replaced during a face-to-face real-world interaction.

Vitevitch (2003) demonstrated that failure to detect changes also occurs in audition. Participants were asked to shadow a list of words that were presented over headphones. Halfway through the list of words, the voice presenting the words was replaced with a different voice. Nearly half of the listeners who were presented with the new voice failed to notice the change. Vitevitch manipulated the lexical difficulty of the shadowed words and measured response times to demonstrate that an interaction of allocating attention and word processing can influence change detection for spoken material. In essence, allocating more attention to word processing increased the likelihood that participants would fail to detect a change in talker.

Subsequent change deafness studies have shown that the acoustic characteristics of the sounds in an auditory scene can influence the detection of auditory change. For example, changes between nonspeech sounds that are acoustically dissimilar are detected more frequently than those between

J. G. Neuhoff (✉) · M. C. Ndiaye · A. B. Berkow ·
B. R. Bertacchi · C. A. Benton
College of Wooster, 1189 Beall Avenue, Wooster, OH 44691, USA
e-mail: jneuhoff@wooster.edu

J. Wayand
Walsh University, North Canton, OH, USA

sounds that are more similar (Gregg & Samuel, 2008; Gregg & Snyder, 2012). The effect also occurs for speech sounds, since changes between male and female voices are detected at higher rates than same-sex changes (Fenn et al., 2011).

In addition to being influenced by the acoustic characteristics of the sound, change deafness has been shown to be modulated by the attention of the listener and the emotional content of the sounds. Some work has even shown that listeners show better change detection in an unfamiliar language than in their native language, despite having better native-language discrimination abilities (Neuhoff, Schott, Kropf, & Neuhoff, 2014). Presumably, processing the semantic information in a familiar language draws attention from the indexical characteristics of the speaker. In an unfamiliar language, semantic information is unavailable, and listeners focus their attention on the acoustic properties of the voice. Other work has shown that changes in negatively valenced emotional stimuli (e.g., a growling dog) are detected at higher rates than emotionally neutral stimuli (Asutay & Vastfjall, 2014).

Change deafness studies have typically used some kind of interruption or pause between the two stimuli used to produce the change. For example, in a telephone-based change deafness study, listeners answered survey questions and were put on hold briefly while a change in speakers occurred (Fenn et al., 2011). Other studies have used periods of silence, bursts of noise, or other interruptions (Constantino, Pinggera, Paranamana, Kashino, & Chait, 2012; Fenn et al., 2011; Gregg, Irsik, & Snyder, 2014; Gregg & Samuel, 2008; Vitevitch, 2003). The use of an interruption between two auditory stimuli in a change deafness experiment is analogous to a visual interruption or mask often employed in visual change blindness experiments. Some researchers have suggested, for example, that a burst of noise presented between stimuli can mask transients and echoic memory traces that might be used to more effectively detect the change (Eramudugolla, Irvine, McAnally, Martin, & Mattingley, 2005). One study has shown that in an auditory scene with multiple sound sources, change deafness without an intervening interruption occurred as often as when the two auditory scenes were separated by 500 ms of silence (Pavani & Turatto, 2008). However, in this study the extent to which the other, simultaneously sounding objects in the scene might have masked the change is unclear. It should also be noted that this study used multiple trials in a within-subjects design with a total of only 12 potential sounds, all drawn from the same semantic category (animal calls). Thus, it is unclear how repeated presentation of the same stimuli both with and without an intervening noise burst might have dissipated the effects of the masking sound.

Although there are undoubtedly important differences between auditory and visual change detection (Demany, Semal, Cazalets, & Pressnitzer, 2010; Demany, Trost, Serman, & Semal, 2008), change blindness studies have also typically presented an analogous mask or distractor between two visual

scenes when a change occurs (for reviews, see Simons, 2000; Simons & Rensink, 2005). The distractor between scenes is thought to draw attention away from the change, thus making it more difficult to detect (O'Regan, Rensink, & Clark, 1999; Simons, Franconeri, & Reimer, 2000). However, some work has shown that change blindness can occur with above-threshold visual changes in the absence of a distractor if the changes occur slowly and continuously over an extended period of time. Simons et al. (2000) presented viewers with complex visual scenes in which a single object gradually appeared or disappeared over 12 s, and they found that participants failed to notice the changes well over 40 % of the time. These results clearly demonstrated that a distractor is not necessary to induce change blindness when the visual change that occurs is gradual and does not create a transient. The results also suggested that the level of detail in a scene that is retained from one viewing to the next is less than had previously been thought.

Work in speech perception has suggested that the indexical characteristics of the speaker are implicitly encoded. For example, previous studies have shown strong priming effects on intelligibility, word shadowing, and stem completion tasks when the training voice in the experiment was the same as the test voice (Goldinger, 1998; Nygaard & Pisoni, 1998; Schacter & Church, 1992). If the voices at training and test are different, performance declines. More recent work has implicated fundamental frequency as a contributing factor to these effects. For example, when a single speaker's voice is artificially shifted in fundamental frequency between training and test, similar declines in performance result (Church & Schacter, 1994). These results suggest an implicit memory for vocal fundamental frequency that may at first appear to be at odds with more recent studies of change deafness: Listeners automatically encode the properties of a voice, yet unless they are alerted to the possibility of a change, they often fail to notice when dramatic changes in that voice occur.

In vision, gradual changes that occur without an intervening transient often go unnoticed (Simons et al., 2000). In the present work, we examined whether a similar phenomenon occurs in audition. We presented listeners with continuous speech stimuli that exhibited a gradual and continuous change in pitch over time and measured the rate of change detection using a standard change deafness paradigm. We found that many listeners failed to detect changes in vocal pitch even when the speech signal was changed up to six semitones.

Experiment 1

Method

Participants Our sample consisted of 80 native English speakers from the United States (mean age = 32.2 (*SD* 11.1);

47 male, 33 female). All reported normal hearing; they were recruited via Amazon Mechanical Turk (MTurk) and were paid \$0.30 to complete the experiment online. A wide variety of research has shown that samples from MTurk are more diverse and more representative of target populations, and that their reliability is as good as or better than that obtained from traditional undergraduate samples (Buhrmester, Kwang, & Gosling, 2011; Crump, McDonnell, & Gureckis, 2013; Holden, Dennie, & Hicks, 2013; Mason & Suri, 2012; Paolacci, Chandler, & Ipeirotis, 2010).

Stimuli We recorded a male and a female voice reading the same passage from *World War Z* by Max Brooks (see the Appendix). The stereo recordings were saved as .wav files, matched for overall RMS power, and submitted to Praat (Boersma & Weenink, 1992) for analysis of the mean fundamental frequency (f_0) of each voice (female, 186.5 Hz; male, 106.5 Hz). The durations of the audio clips were 104.3 s (female) and 102.6 s (male). We then used editing software (CoolEdit Pro, Syntrillium Software) to create “rising”- and “falling”-pitch versions of each clip. The rising-pitch versions started 1.5 semitones below the actual voice pitch, rose gradually and continuously over the duration of the clip, and ended 1.5 semitones above the actual pitch of each voice, yielding a three-semitone change in pitch from beginning to end. The falling-pitch versions started 1.5 semitones above the actual voice pitch and ended 1.5 semitones below the actual voice pitch (see Fig. 1). The three-semitone pitch modification was applied uniformly to the entire signal and was accomplished without changing the rates or durations of the audio clips. The four resulting stimuli were male and female versions of the rising- and falling-pitch conditions.

Design and procedure Participants were randomly assigned to one of the four Speaker Sex \times Direction of Pitch Change conditions in a completely between-subjects design.

Participants were instructed to listen to the story and count the number of breaths taken by the speaker. Prior to beginning the experiment, the participants were instructed to adjust their listening volume to a comfortable level while listening to an instruction audio file recorded in a synthetic text-to-speech voice. As a check that participants could actually hear the audio, they were also instructed to enter a code word that appeared at the end of the introductory audio clip. The synthetic voice said: *Please take a moment to adjust your listening volume to a comfortable level. The word that you should type onto the space below is “apple.”* Three of the 80 participants failed to enter the correct code word and were replaced. At the conclusion of the story, all participants were asked three change deafness questions, adapted from Vitevitch (2003): (1) *Did you notice anything unusual about the experiment?*, (2) *Did you notice any changes in the speaker’s voice?*, and (3) *Did the voice sound the same at the beginning of the passage as it did at the end?* Question 3 responses were used as the primary unit of analysis. The first question was open ended, and the next two were yes/no. However, to follow up on positive change detection responses, those participants who responded “yes” to Question 2 were then asked *What changes did you notice?* Those responding “yes” to Question 3 were subsequently asked *How was the voice different at the beginning and the end?*

Results and discussion

In an analysis of change detection, we found that only 42/80 participants (52.5 %) noticed the three-semitone change in vocal pitch over the duration of the story, either by indicating in response to Question 1 that the pitch of the voice had changed or by answering “yes” to Question 2 or “no” to Question 3. The remaining 38/80 participants (47.5 %) failed to notice the change and indicated that the voice sounded the same at the beginning and the end of the passage. Neither

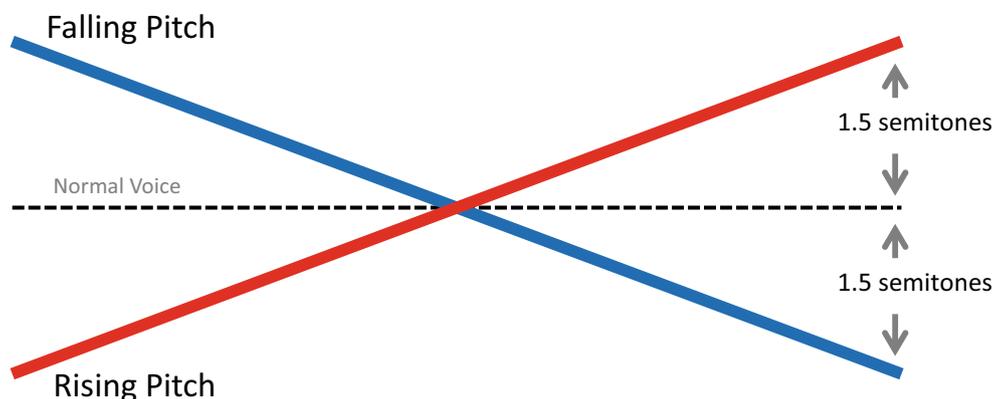


Fig. 1 Change in vocal pitch employed for the stimuli in Experiment 1. Rising-pitch versions started 1.5 semitones below the actual voice pitch, rose over the duration of the clip, and ended 1.5 semitones above the

actual pitch of each voice. Falling-pitch versions started 1.5 semitones above and ended 1.5 semitones below the actual voice pitch.

talker sex, $\chi^2(1) = 0.50$, $p = .82$, nor the direction of pitch change, $\chi^2(1) = 1.3$, $p = .26$, significantly influenced the detection of change.

The 42 participants who indicated that the voice had changed were also asked *How was the voice different at the beginning and the end?* Of these participants, 36/42 (85.7 %) submitted responses indicating a detection of the pitch change that was implemented. Most of these responses directly mentioned pitch change, with a few indicating that the end voice was either a different person or a younger/older voice. The six remaining responses indicated that the voice at the end was “faster” (2), “more upbeat,” “weaker,” “more nervous,” and “a little looser.” Thus, actual change deafness for the stimulus characteristics that were manipulated might even be slightly higher than our conservatively measured 47.5 %. Although some of these responses may be indicative of a detection of pitch change that was not articulated well by the participants, they do suggest that a 47.5 % change deafness rate under these conditions may be a conservative estimate. We found no differences in the numbers of breaths counted between those that noticed the change ($M = 20.6$, $SD = 9.1$) and those that failed to notice the change ($M = 22.0$, $SD = 14.3$), $t(78) = 0.52$, $p = .61$.

The results of Experiment 1 show that change deafness can occur for gradually and continuously changing stimuli without a masking or intervening silent interval to disrupt detection of the change. This finding is consistent with previous work that had used auditory scenes composed of multiple sound sources and shown that change deafness for discrete changes can occur without an intervening masking interval (Pavani & Turatto, 2008). Although the mechanisms of change detection may be different in a multisource listening environment, the present results suggest that change deafness for a single isolated sound source can also occur without a discrete break or interruption occurring to mask the change, if the change takes place gradually over time.

A similar “slow-change” blindness effect has been demonstrated in vision. Simons et al. (2000) showed that observers failed to notice both gradual changes in the color of, and the slow appearance or disappearance of, objects in a complex visual scene. The failure to detect slow changes in vision has been interpreted as supporting the more general change blindness hypothesis that only a small portion of the details from a visual scene are retained and cognitively represented from one view of the scene to the next. The present results suggest that a similar phenomenon may occur with respect to the representation of indexical characteristics of the voice.

However, another potential explanation for these results could be that our sample simply lacked the sensitivity to perform the task, because accuracy rates in our two-alternative forced choice task were near 50 %. Although the high degree of accuracy in the follow-up question asking *what* change occurred suggests otherwise, it could be that the change over

time was simply too small to detect. In addition, participants in the present study performed a concurrent distractor task of counting the number of breaths taken by the speaker. Vitevitch (2003) used a concurrent verbal shadowing task in a similar manner and found that participants who noticed a change in speaker performed more poorly on the shadowing task when the words were difficult. Thus, the effect of the concurrent “breath counting” task on the detection of vocal pitch change might have contributed to the high rates of change deafness for gradual vocal pitch changes. We examined both of these issues in Experiment 2.

Experiment 2

In Experiment 1, nearly 50 % of our participants failed to notice a three-semitone change in vocal pitch while they were asked to count the breaths taken by the speaker. Experiment 2 was designed to examine the degree to which this concurrent task might have contributed to the change deafness rate for pitch change in Experiment 1. We also wanted to examine the discriminability of the three-semitone change in pitch over the duration of the story that was used in Experiment 1. It is possible that a three-semitone change over the average 103.5-s duration was at the sensory limit for detecting changes in vocal pitch. It may also be that the task was at the limits of short-term memory capabilities for vocal pitch. For example, one way to successfully detect the change would be to compare the voice at the end of the story to a memory of the voice at the beginning. If this task exceeded the capacity of short-term memory for many listeners, then our “change deafness” results in Experiment 1 would simply be due to the task difficulty.

To examine these issues, we presented two groups of listeners with the stimuli that had been used in Experiment 1, but gave each group a different set of instructions. To determine whether most listeners would detect this type of pitch change if alerted to it beforehand, we told one group to listen for a gradual change in pitch over the duration of the story. If the change deafness observed in Experiment 1 was simply a function of the change being at or near threshold or of the limitations of auditory memory for vocal pitch, then in this condition we should expect a rate of change deafness similar to that found in Experiment 1. On the other hand, if the change deafness exhibited in Experiment 1 occurred because of factors typically implicated in other change deafness investigations (e.g., failures of attention, cognitive representation, comparison, etc.), then we would expect to find much lower rates of change deafness when listeners were alerted to listen for such changes (Backer & Alain, 2012; Eramudugolla et al., 2005). The second group of participants was instructed

simply to listen to the story and to be prepared to answer some questions when it concluded. This condition simultaneously provided an evaluation of the effect of counting breaths in Experiment 1 and a control condition for the present experiment. If the concurrent task of counting breaths in Experiment 1 inflated the change deafness rates, we might expect to find lower rates of change deafness in this condition.

Method

Participants A total of 160 native English speakers from the United States (mean age = 33.1 (*SD* 11.3); 86 male, 74 female) were recruited via MTurk and were paid \$0.30 to complete the experiment online. All reported normal hearing. None had participated in the previous experiment.

Stimuli The stimuli were the same as those used in Experiment 1.

Design and procedure Participants were instructed to adjust their listening volume to a comfortable level and to enter a code word, as in Experiment 1. Five participants failed to enter the correct code word and were replaced. At the conclusion of the story, all were asked the three initial and two follow-up change deafness questions from Experiment 1. In addition, we asked a three-alternative forced choice question: (4) *Did the voice pitch change up, down, or stay the same?*

“Listen for pitch change” instructions Half of the participants were directly instructed to listen to the story and to determine whether the pitch of the voice changed over the duration of the clip. Participants in this condition were told not to be concerned with the natural up and down changes in pitch that occur as people speak naturally, but to listen for a gradual change in pitch over the duration of the story, as if the speaker’s voice sounded higher or lower or older or younger from the beginning to the end. They were also told to be prepared for some questions at the end of the story.

“Just listen” instructions The remaining participants were simply instructed to listen to the story and to be prepared for some questions at the end.

Results and discussion

Change detection In the “listen for pitch change” condition, 67/80 participants (83.75 %) noticed the change, by indicating in response to either Question 1 or 2 that the pitch of the voice had changed or by answering

“no” to Question 3. However, when participants were simply told to “just listen,” only 43/80 (53.75 %) detected the change in pitch (see Fig. 2 for these results, graphed as “% change deafness”). The difference between these two conditions was significant, $\chi^2(1) = 12.20$, $p = .0005$. These results suggest that the gradual changes in vocal pitch that we introduced were easily detected when listeners were alerted to the possibility of a change. Thus, the task does not appear to exceed the sensory limits or capabilities of auditory short-term memory. However, when not explicitly listening for a change, nearly half of our participants failed to detect it.

These findings are consistent with previous work that had shown that orienting attention to potential auditory changes can reduce change deafness (Backer & Alain, 2012; Eramudugolla et al., 2005). That we found roughly the same proportion of change deafness for listeners in the “just listen” condition as in the “count breaths” condition of Experiment 1 suggests that the breath-counting task in Experiment 1 was likely not a factor that contributed to change deafness for gradually changing voice pitch.

Previous work on detecting slow changes in visual stimuli has shown that when specifically directed to look for objects gradually entering or fading from a complex visual scene, participants detected the change 64.3 % of the time (Simons et al., 2000). When instructed to listen for gradual change, our participants detected the change 83.75 % of the time. However, in addition to the modality differences, other important differences between these two experiments make direct comparison difficult. Simons et al. presented a change in one object among many in an otherwise static visual scene. Here we presented a single voice in an isolated auditory scene and changed one perceptual dimension of the voice. The

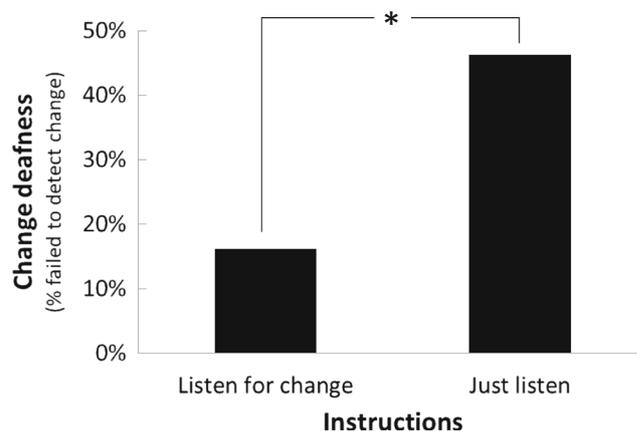


Fig. 2 Results from Experiment 2: Listeners who were instructed to listen for a gradual change in vocal pitch exhibited significantly less change deafness than those instructed simply to listen to the voice. * $p < .05$

duration of the change was also much longer in the present study. Future work might examine more equally matched conditions in order to investigate modality differences in slow change detection.

Direction of pitch change detection We next examined accuracy rates for detecting the *direction* of pitch change. We asked all participants a three-alternative forced choice question: (4) *Did the voice pitch change up, down, or stay the same?*

Failed to notice change For participants in the “just listen” condition who failed to notice the change, 25/37 (67.5 %) reported that the pitch stayed the same. Of the remaining 12 responses, which indicated that a change had occurred, 5/12 (41.6 %) incorrectly reported the direction of change. For the 13 participants who failed to notice the change in the “listen for change” condition, 9/13 (69.2 %) reported that the pitch stayed the same. Of the remaining participants, 4/4 (100 %) were correct in indicating the direction of change.

Noticed change For the 43 participants in the “just listen” condition who indicated that they had noticed a change in pitch, 11/43 (25.6 %) participants were incorrect in identifying the direction of the change. Thus, if we combine those who were incorrect in detecting the direction of pitch change with those who failed to notice, we find that 48/80 (60 %) failed to correctly detect the direction of pitch change. When we conducted a similar analysis on the “listen for change” group, we found that a significantly smaller proportion of those who noticed a change (4/67, 6 %) were incorrect in identifying the direction of pitch change, $\chi^2(1) = 6.97$, $p = .008$ (Yates correction). Thus, in the directed-attention condition, not only were participants more likely to detect a change, but those who did detect a change were significantly more accurate in detecting the direction of that change. Other studies have also shown that listeners sometimes can accurately detect pitch change but fail to identify the direction of the change (Mathias, Micheyl, & Bailey, 2010; Neuhoff, Knight, & Wayand, 2002; Semal & Demany, 2006). However, the duration of the stimuli used in the previous work was considerably shorter than that used in the present experiment.

Experiment 3

In Experiment 3, we assessed the discriminability of the vocal pitch at the beginning versus the end of the passages used in

the previous experiments. Experiment 2 showed that when they were directed to listen for a change over the course of the passage, over 80 % of participants detected the change. Here we presented listeners with an AX discrimination task in which they made same–different judgments about the pitch of the first and last sentences from the stimuli used in the previous experiments. Listeners heard the first sentence transposed either up or down 1.5 semitones from the original voice, followed by the last sentence either at the same pitch or transposed 1.5 semitones from the original in the opposite direction. This allowed us to assess the discriminability of the three-semitone change from the beginning to the end of the passage in a quick and discrete manner, rather than the slow, gradual change that had taken place in Experiment 2.

Method

Participants A group of 80 native English speakers from the United States (mean age = 28.0 (SD 9.3); 52 male, 28 female) were recruited via MTurk and were paid \$0.30 to complete the experiment online. All reported normal hearing, and none had participated in the previous experiments.

Stimuli, design, and procedure In an AX within-subjects discrimination task, listeners made same–different judgments about eight pairs of single sentences derived from the recordings used in Experiment 1. Listeners heard the first and last sentences of the narrative used in Experiment 1, separated by 1,000 ms (with a 50-ms chirp in the middle of the interstimulus interval to delineate the two separate stimuli). On half of the trials, listeners heard two sentences of different vocal pitches (low–high or high–low). On the other half of the trials, the vocal pitch was the same (low–low or high–high). In this context, “high” and “low” indicate stimuli transposed up or down 1.5 semitones, respectively, from the original recording. Thus, on “different” trials, listeners heard two sentences separated in vocal pitch by three semitones. On “same” trials, listeners heard two sentences that were the same pitch, both having been transposed 1.5 semitones in the same direction. Participants heard one of each type of trial from both the male and female voices, for a total of eight trials presented in random order. We intentionally kept the number of trials low and the number of participants high (as compared to more traditional AX discrimination experiments), in order to more closely match the conditions in Experiments 1 and 2 (in which a large number of participants were tested and only one trial was possible). After each stimulus pair, listeners indicated whether the voices were the same or different. Prior to beginning the experiment, all participants were instructed to adjust their listening volume to a comfortable level and to enter a code word as in Experiment 1. Two participants failed to enter the correct code word and were replaced.

Results and discussion

The mean proportion of correct responses across all participants was .95 ($SD = .09$). Because the number of trials was small and many participants showed perfect performance, we averaged hits and false alarms across participants and calculated a measure of group sensitivity (following Macmillan & Kaplan, 1985). We found that the stimuli were highly discriminable, with $d' = 3.2$ ($C = 0.04$). We also compared the percentage of participants with perfect change detection in the present experiment (a hit rate equal to 1.0 over the four “different” trials) to the percentage of participants who noticed the continuous change in pitch when directed to listen for it in Experiment 2. In the discrete-change condition, 90 % of the 80 participants always noticed the difference in vocal pitch. In the continuous-change condition, when listeners were instructed to listen for changes, 83.75 % detected the change. These rates of change detection were not significantly different, $z = 1.71$, $p = .24$. Though the conditions under which judgments were made in Experiments 2 and 3 were quite different, the results of both suggest that the changes were well above threshold and that when participants were alerted to the possibility of a change, changes were detected at a very high rate. Thus, the change deafness exhibited when listeners were not alerted to the potential for change did not stem from a fundamental lack of sensitivity to the amount of stimulus change introduced.

Experiment 4

In Experiment 4, we examined how the amount of slow pitch change that occurred in a voice would affect the likelihood that listeners would detect the change. As acoustic differences between sounds become greater, listeners detect changes at a higher rate (Gregg & Snyder, 2012). For example, Fenn et al. (2011) showed that a change from a male to a female voice is detected more often than a same-sex change. In Experiments 1–3, we presented a change in vocal pitch of three semitones. In Experiment 4, we increased this change and presented listeners with either four- or six-semitone changes.

However, because of the fixed vocal tract length, a human voice transformed beyond about three semitones begins to take on “cartoonish” characteristics and to no longer sound plausibly human, without additional modifications based on complex vocal tract and glottal source modeling (Rose, 2009; Türk & Arslan, 2006). Thus, we used a high-quality text-to-speech synthetic voice that retained its essential vocal characteristics throughout the range of pitch transformations. Our goal was to replicate the change deafness phenomena demonstrated in Experiment 1 with different materials and to test the

hypothesis that larger changes in pitch would lead to a lower incidence of change deafness.

Method

Participants A group of 80 native English speakers from the United States (mean age = 30.7 (SD 11.5); 43 male, 37 female) were recruited via MTurk and were paid \$0.30 to complete the experiment online. All reported normal hearing, and none had participated in the previous experiments.

Stimuli We used a high-quality commercial text-to-speech reader with a female American English voice (“Salli”; Ivona Text-to-Speech Software) to render a recording of an excerpt from the story “The Birthday” by Mike McCormick (see the Appendix). The stereo recording was saved as a .wav file and submitted to Praat (Boersma & Weenink, 1992) for analysis of its mean fundamental frequency ($f_0 = 188.4$ Hz). The duration of the audio clip was 124.4 s. We then used CoolEdit Pro (Syntrillium Software) to create “large-change” and “small-change” rising- and falling-pitch versions of the clips. The “large-change” version started three semitones above/below the actual voice pitch and fell/rose gradually and continuously over the duration of the clip, ending three semitones below/above the actual pitch of the voice, yielding a six-semitone rising or falling change in pitch from beginning to end. The “small-change” version started two semitones above/below the actual voice pitch and fell/rose gradually and continuously over the duration of the clip, ending two semitones below/above the actual pitch of the voice, yielding a four-semitone rising or falling change in pitch from beginning to end. Pitch modification was applied uniformly to the entire signal and was accomplished without changing the rate or duration of the audio clips. The four resulting stimuli were four-semitone change and six-semitone change versions in both the rising- and falling-pitch conditions.

Design and procedure Half of the stimuli changed four semitones, and half changed six; half were rising in pitch, and half were falling. Listeners were randomly assigned to one of the four conditions and were instructed to listen to the story and count the number of “natural pauses” that occurred. Because a computerized voice does not actually “breathe,” participants were instructed to count the number of natural pauses that occurred, where a real person might take a breath. Prior to beginning the experiment, all participants were instructed to adjust their listening volume to a comfortable level and to enter a code word, as in Experiment 1. Three participants failed to enter the correct code word and were replaced. At the conclusion of the story, all participants were asked the same set of questions used in Experiment 1.

Results and discussion

We found that across both pitch change conditions, 49 of our 80 participants (61.25 %) detected a change in vocal pitch over the duration of the narrative, indicated by mentioning that the voice changed in response to any of the postexperiment questions. We examined the effect of the amount of pitch change on the tendency to detect a change and found that significantly more participants detected the six-semitone change (29/40, 72.5 %) than detected the four-semitone change (20/40, 50 %), $\chi^2(1) = 4.3$, $p = .039$ (see Fig. 3 for these results, graphed as “% change deafness”).

The 20 participants who did notice that the voice was different in the four-semitone change condition were also asked the follow-up question *How was the voice different at the beginning and the end?* Of these participants, 16/20 submitted responses indicating detection of the actual change that had been implemented. Twelve of these responses directly mentioned pitch change, with four indicating that the end voice was either a different person or a younger/older voice. The remaining responses indicated that the voice at the end was more “robotic,” “steady,” “upset,” and “soft.”

The 29 participants who detected a change in the six-semitone condition were also asked the same follow-up question. Of these, 25 submitted responses that addressed the change in pitch. The remaining four indicated that the voice became “more robotic,” “more computerized,” “slower,” and “had fewer pauses.” Thus, actual change deafness for the stimulus characteristics that were manipulated might be slightly higher than the original analysis suggested.

These results replicate the finding that gradual changes in vocal pitch that would otherwise be very noticeable can go undetected if listeners are not alerted to the possibility of a change. We also found support for previous work that had shown that increasing the amount of an acoustic change will decrease the rate of change deafness (Fenn et al., 2011; Gregg

& Snyder, 2012). The latter finding highlights the contributions of both sensory and attentional factors in detecting auditory change.

General discussion

In four experiments, we have shown that, as in a similar phenomenon in vision, listeners often fail to detect gradual and continuous changes in auditory stimuli. In Experiment 1, we presented listeners with continuous speech that changed three semitones in pitch over time, and found that nearly 50 % failed to notice the change. Experiments 2 and 3 demonstrated that the changes in the stimuli were well above threshold and that when listeners were alerted to the possibility of a change, detection rates improved dramatically. These results are consistent with previous work that had shown that cueing listeners to potential auditory changes can significantly reduce change deafness (Alain & Bernstein, 2008; Backer & Alain, 2012; Vitevitch, 2003). Experiment 4 provided support for previous work that had shown that increasing the magnitude of a change in the stimulus will decrease the rate of change deafness (Fenn et al., 2011; Gregg & Snyder, 2012).

Our results showed that failure to detect talker changes extends to stimuli that change gradually over time. Previous work had suggested that the indexical characteristics of a speaker’s voice are automatically encoded, are retained in long-term memory, and can influence subsequent performance on linguistic tasks that do not necessarily depend on the specific indexical characteristic learned (Goldinger, 1998; Nygaard & Pisoni, 1998; Nygaard, Sommers, & Pisoni, 1994; Palmeri, Goldinger, & Pisoni, 1993; Schacter & Church, 1992). Yet despite these findings, change blindness studies have routinely shown that listeners can fail to notice when a talker changes. The present results show that this failure to detect change occurs not only with discrete changes between talkers, but also when a single voice changes gradually and continuously.

Magnuson and Nusbaum (2007) demonstrated that when listeners expect to hear two different talkers, performance on a speeded target-monitoring task decreases relative to listeners who hear the same stimuli but expect only one talker. The performance deficit is thought to result from the additional cognitive resources required to adjust to talker variability, and the phenomena has been shown to be directly related to auditory change deafness (Vitevitch, 2003). Thus, when listeners are aware of the possibility of a change, they can devote resources to monitoring for it. With no expectation of change, these resources can be better spent processing the acoustic scene.

Our results also demonstrate a phenomenon that in some ways is analogous to the failure to detect slow changes in vision (Simons et al., 2000). However, there are some key

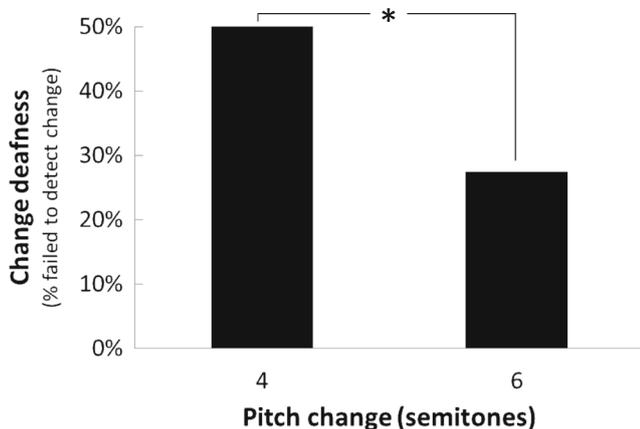


Fig. 3 Results from Experiment 4: Listeners exhibited significantly greater change deafness when the change in pitch was four rather than six semitones. * $p < .05$

differences. In vision, slow-change blindness has so far been demonstrated with a change in one attribute of an otherwise static but complex visual scene. In the present work, we presented listeners with a single auditory object that changed gradually. The presence of visual objects in the scene that are *not* changing may make the slow change that is occurring harder to detect. Future work might explore a closer auditory analogue in which one sound in a complex auditory scene exhibits a slow change, or even explore multimodal change detection, as has been examined in inattention blindness (Wayand, Levin, & Varakin, 2005).

Conclusions

We have demonstrated a new phenomenon whereby listeners fail to detect large gradual changes in the pitch of a voice when no change is expected. Our results are consistent with an account of change deafness that takes into account both the magnitude of a stimulus change and listener expectations.

Appendix

Story from Experiments 1 and 2

The sky was red that day. All the smoke, the crap that'd been filling the air all summer. It put everything in an amber red light, like looking at the world through hell-colored glasses. That's how I first saw Yonkers, this little, depressed, rust-collar burb just north of New York City. I don't think anybody ever heard of it. I sure as hell hadn't, and now it's up there with, like, Pearl Harbor . . . no, not Pearl . . . that was a surprise attack. This was more like Little Bighorn, where we . . . well . . . at least the people in charge, they knew what was up, or they should have. The point is, it wasn't a surprise, the war . . . or emergency, or whatever you want to call it . . . it was already on. It had been, what, three months since everyone jumped on the panic train.

You remember what it was like, people just freaking out . . . boarding up their houses, stealing food, guns, shooting everything that moved. They probably killed more people, the Rambos and the runaway fires, and the traffic accidents and just the . . . the whole shit storm that we now call "the Great Panic"; I think that killed more people at first than Zack.

I guess I can see why the powers that be thought that one big stand-up battle was such a good idea. They wanted to show the people that they were still in charge, get them to calm the hell down so they could deal with the

real problem. I get it, and because they needed a propaganda smackdown, I ended up in Yonkers.

It actually wasn't the worst place to make a stand. Part of the town sat right in this little valley, and right over the west hills you had the Hudson River. The Saw Mill River Parkway ran right through the center of our main line of defense and the refugees streaming down the freeway were leading the dead right to us. It was a natural choke point, and it was a good idea . . . the only good idea that day.

Story text from Experiment 4

Dad was carrying a tray of grilled hot dogs across the fresh-cut grass when his knees stiffened and black smoke spewed from his nose. His shoulders lurched forward and his chin swung down and clanged against his chest as more and more black smoke billowed from his open mouth. There was a loud bang and a child's scream as Dad crashed to the ground face-down, elbows pointed to the sky, his body still locked in the position of a man carrying a tray of hot dogs.

Mom herded the screaming kids into the kitchen and called my Uncle Steve who was there within ten minutes. In the meantime she pushed red candles into my cake and led a cheery chorus of "Happy Birthday" over the bangs that sounded from the backyard.

Uncle Steve came and ate a piece of my vanilla cake with his bare hands. He hugged Mom and leaned down and said Happy Birthday to me.

"Want to help me with your Pop for a moment?" he said. Mom smiled and hung a cardboard donkey on the pantry door. Uncle Steve held the back door open and I walked through.

Dad was still smoking when we got to him. Sometimes a spark would flash in the air around his head. Uncle Steve gave me a metal box to hold and bent down to peer into Dad's ear canal.

"Should we touch him?" I said.

"We're going to have to" said Uncle Steve.

He went to work with his screwdriver, pliers, and some other tools I hadn't seen before. It was only my tenth birthday. I didn't know much about tools. I didn't know much about Uncle Steve either, except that I thought he could probably read minds.

He banged around for a while, tugging at Dad's arms and wrenching at something under his armpit. Then he leveraged a small instrument down Dad's throat. He placed his finger between Dad's teeth, feeling for something. Then he pulled and we heard a loud click.

"Aha!" he said.

Dad wasn't smoking anymore. Uncle Steve knocked on the back of Dad's neck, listening for a certain sound. The air around us began to clear. I could hear laughter coming from my house.

References

- Alain, C., & Bernstein, L. (2008). From sounds to meaning: The role of attention during auditory scene analysis. *Current Opinion in Otolaryngology and Head and Neck Surgery*, *16*, 485–489. doi:10.1097/MOO.0b013e32830e2096|10.1097/MOO.0b013e32830e2096
- Asutay, E., & Vastfjall, D. (2014). Emotional bias in change deafness in multisource auditory environments. *Journal of Experimental Psychology: General*, *143*, 27–32. doi:10.1037/a0032791
- Backer, K. C., & Alain, C. (2012). Orienting attention to sound object representations attenuates change deafness. *Journal of Experimental Psychology: Human Perception and Performance*, *38*, 1554–1566. doi:10.1037/a0027858
- Boersma, P., & Weenink, D. (1992). Praat: Doing phonetics by computer [Computer program]. Retrieved from www.fon.hum.uva.nl/praat/
- Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, *6*, 3–5. doi:10.1177/1745691610393980
- Church, B. A., & Schacter, D. L. (1994). Perceptual specificity of auditory priming: Implicit memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 521–533. doi:10.1037/0278-7393.20.3.521
- Constantino, F. C., Pinggera, L., Paranamana, S., Kashino, M., & Chait, M. (2012). Detection of appearing and disappearing objects in complex acoustic scenes. *PLoS ONE*, *7*, e46167. doi:10.1371/journal.pone.0046167
- Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLoS ONE*, *8*, e57410. doi:10.1371/journal.pone.0057410
- Demany, L., Semal, C., Cazalets, J., & Pressnitzer, D. (2010). Fundamental differences in change detection between vision and audition. *Experimental Brain Research*, *203*, 261–270. doi:10.1007/s00221-010-2226-2
- Demany, L., Trost, W., Serman, M., & Semal, C. (2008). Auditory change detection—Simple sounds are not memorized better than complex sounds. *Psychological Science*, *19*, 85–91. doi:10.1111/j.1467-9280.2008.02050.x
- Eramudugolla, R., Irvine, D. R. F., McAnally, K. I., Martin, R. L., & Mattingley, J. B. (2005). Directed attention eliminates “change deafness” in complex auditory scenes. *Current Biology*, *15*, 1108–1113. doi:10.1016/j.cub.2005.05.051
- Fenn, K. M., Shintel, H., Atkins, A. S., Skipper, J. I., Bond, V. C., & Nusbaum, H. C. (2011). When less is heard than meets the ear: Change deafness in a telephone conversation. *Quarterly Journal of Experimental Psychology*, *64*, 1442–1456. doi:10.1080/17470218.2011.570353
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*, 251–279. doi:10.1037/0033-295X.105.2.251
- Gregg, M. K., Irsik, V. C., & Snyder, J. S. (2014). Change deafness and object encoding with recognizable and unrecognizable sounds. *Neuropsychologia*, *61*, 19–30. doi:10.1016/j.neuropsychologia.2014.06.007
- 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. doi:10.1037/0096-1523.34.4.974
- Gregg, M. K., & Snyder, J. S. (2012). Enhanced sensory processing accompanies successful detection of change for real-world sounds. *NeuroImage*, *62*, 113–119. doi:10.1016/j.neuroimage.2012.04.057
- Grimes, J. (1996). On the failure to detect changes in scenes across saccades. In K. Akins (Ed.), *Perception* (Vol. 5, pp. 89–110). Oxford: Oxford University Press.
- Holden, C. J., Dennie, T., & Hicks, A. D. (2013). Assessing the reliability of the M5-120 on Amazon's mechanical Turk. *Computers in Human Behavior*, *29*, 1749–1754. doi:10.1016/j.chb.2013.02.020
- 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*, 185–199. doi:10.1037/0033-2909.98.1.185
- 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. doi:10.1037/0096-1523.33.2.391
- Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, *44*, 1–23. doi:10.3758/s13428-011-0124-6
- Mathias, S. R., Micheyl, C., & Bailey, P. J. (2010). Stimulus uncertainty and insensitivity to pitch-change direction. *Journal of the Acoustical Society of America*, *127*, 3026–3037. doi:10.1121/1.3365252
- Neuhoff, J. G., Schott, S. A., Kropf, A. J., & Neuhoff, E. M. (2014). Familiarity, expertise, and change detection: Change deafness is worse in your native language. *Perception*, *43*, 219–222. doi:10.1068/p7665
- Neuhoff, J. G., Knight, R., & Wayand, J. (2002). *Pitch change, sonification, and musical expertise: Which way is up?* Article presented at the International Conference on Auditory Display, Kyoto, Japan.
- Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, *60*, 355–376. doi:10.3758/bf03206860
- Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, *5*, 42–46. doi:10.1111/j.1467-9280.1994.tb00612.x
- O'Regan, J. K., Rensink, R. A., & Clark, J. J. (1999). Change-blindness as a result of “mudsplashes”. *Nature*, *398*, 34. doi:10.1038/17953
- 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*, 309–328. doi:10.1037/0278-7393.19.2.309
- Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, *5*, 411–419.
- Pavani, F., & Turatto, M. (2008). Change perception in complex auditory scenes. *Perception & Psychophysics*, *70*, 619–629. doi:10.3758/PP.70.4.619
- Rensink, R. A. (2002). Change detection. *Annual Review of Psychology*, *53*, 245–277. doi:10.1146/annurev.psych.53.100901.135125
- Rensink, R. A., Oregan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, *8*, 368–373. doi:10.1111/j.1467-9280.1997.tb00427.x
- Rose, J. (2009). *Audio postproduction for film and video*. Burlington: Elsevier.
- 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*, 915–930. doi:10.1037/0278-7393.18.5.915

- Semal, C., & Demany, L. (2006). Individual differences in the sensitivity to pitch direction. *Journal of the Acoustical Society of America*, *120*, 3907–3915. doi:[10.1121/1.2357708](https://doi.org/10.1121/1.2357708)
- Simons, D. J. (2000). Current approaches to change blindness. *Visual Cognition*, *7*, 1–15. doi:[10.1080/135062800394658](https://doi.org/10.1080/135062800394658)
- Simons, D. J., Franconeri, S. L., & Reimer, R. L. (2000). Change blindness in the absence of a visual disruption. *Perception*, *29*, 1143–1154. doi:[10.1068/p3104](https://doi.org/10.1068/p3104)
- Simons, D. J., & Levin, D. T. (1998). Failure to detect changes to people during a real-world interaction. *Psychonomic Bulletin & Review*, *5*, 644–649. doi:[10.3758/BF03208840](https://doi.org/10.3758/BF03208840)
- Simons, D. J., & Rensink, R. A. (2005). Change blindness: Past, present, and future. *Trends in Cognitive Sciences*, *9*, 16–20. doi:[10.1016/j.tics.2004.11.006](https://doi.org/10.1016/j.tics.2004.11.006)
- Türk, O., & Arslan, L. M. (2006). Robust processing techniques for voice conversion. *Computer Speech and Language*, *20*, 441–467.
- Vitevitch, M. S. (2003). Change deafness: The inability to detect changes between two voices. *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 333–342. doi:[10.1037/0096-1523.29.2.333](https://doi.org/10.1037/0096-1523.29.2.333)
- Wayand, J. F., Levin, D. T., & Varakin, A. (2005). Inattention blindness for a noxious multimodal stimulus. *American Journal of Psychology*, *118*, 339–352.