Identification of familiar voices in disguised speech  

Jessica Clark, Paul Foulkes  
Department of Language and Linguistic Science, University of York, UK  
jhc110@york.ac.uk, pf11@york.ac.uk

Disguise can take many forms, and can be very damaging to both lay and technical speaker identification. We report an experimental study into the ability of lay listeners to identify disguised familiar voices. Similar studies have been performed before, for example by Hollien, Majewski and Doherty (1982). They report relatively good identification rates by lay listeners, but the speech samples in their tests used various forms of disguise including changes to pitch and regional accent. Listeners also varied in the degree of familiarity they had with the tested voices.

In our experiments we tested the effects of pitch modification as a disguise. Pitch was chosen as the parameter for investigation because it can be manipulated in almost all commercially-available voice changing devices. To control for listeners' degree of familiarity with the voices we trained listeners to identify a group of four male speakers. These speakers were selected from the Leeds recordings of the IViE corpus (Grabe, Post and Nolan 2001). Training data consisted of samples of approximately 90 seconds, extracted from readings of the Cinderella story to control the spoken material. 36 listeners were recruited to participate in the experiments. Listeners were aged 19–55 and from a variety of backgrounds, but all resident in York. They were first trained to identify the four speakers from the 90 second samples. They were then asked to listen to 8 experimental stimuli, which consisted of 10 second samples extracted from other sections of the story. They were given a closed test in which the names of the four potential speakers were listed. This test is referred to as the Control condition. The average correct score in the Control condition was 4.78 (out of 8). Individual scores ranged from 8 to 0 (in one case), with a mode of 6.

Four experimental conditions were created in which the voice samples were disguised by F0 modification. Instead of using a particular voice changing device we simulated the effects of pitch modification using the Sound Forge software system in order to control precisely the pitch modification. The experimental stimuli from the Control condition were modified so as to change F0 without changing duration. The four conditions were +8, +4, –4 and –8 semitones. These +8 and –8 values were chosen as the maximum values for modification before the resultant stimuli became unnatural sounding. The listeners were then asked to perform a similar identification test as in the Normal condition. They were split into two groups of 18 for this purpose, one group hearing the +8 and –4 semitone stimuli, and the other group the +4 and –8 semitone stimuli. Training was given in the same mode as in the Control condition. The identification tests again consisted of 8 stimuli in a closed test.

The results of the disguised conditions showed, as predicted, that identification rates fall when listeners hear disguised voices. The average scores were 3.47 in the +8 condition, 3.72 (+4), 3.65 (–4) and 2.50 (–8). The lower success rate was statistically significant in all cases. However, the results for the stimuli in which pitch was raised were better than those in which pitch was lowered. More significant, perhaps, was the variation in performance of individual listeners. In the +4 and –4 conditions the best listeners scored 7/8, while in the +8 and –8 conditions the best score was 5/8. The majority of listeners performed at a level better than chance in all conditions except –8. This suggests that, although disguised speech can be damaging for identification it need not be wholly catastrophic, at least for some listeners.

References