A “voice inversion effect?”

Catherine Bédard and Pascal Belin∗

Département de Psychologie, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal, Que., Canada H3C 3J7

Accepted 12 February 2004
Available online 2 April 2004

Abstract

Voice is the carrier of speech but is also an “auditory face” rich in information on the speaker’s identity and affective state. Three experiments explored the possibility of a “voice inversion effect,” by analogy to the classical “face inversion effect,” which could support the hypothesis of a voice-specific module. Experiment 1 consisted of a gender identification task on two syllables pronounced by 90 speakers (boys, girls, men, and women). Experiment 2 consisted of a speaker discrimination task on pairs of syllables (8 men and 8 women). Experiment 3 consisted of an instrument discrimination task on pairs of melodies (8 string and 8 wind instruments). In all three experiments, stimuli were presented in 4 conditions: (1) no inversion; (2) temporal inversion (e.g., backwards speech); (3) frequency inversion centered around 4000 Hz; and (4) around 2500 Hz. Results indicated a significant decrease in performance caused by sound inversion, with a much stronger effect for frequency than for temporal inversion. Interestingly, although frequency inversion markedly affected timbre for both voices and instruments, subjects’ performance was still above chance. However, performance at instrument discrimination was much higher than for voices, preventing comparison of inversion effects for voices vs. non-vocal stimuli. Additional experiments will be necessary to conclude on the existence of a possible “voice inversion effect.”

© 2004 Elsevier Inc. All rights reserved.

Keywords: Face perception; Face inversion; Recognition; Auditory processing; Object recognition; Voice perception

1. Introduction

The “face inversion effect” is the observation that picture–plane inversion impairs face recognition to a greater extent than object recognition (Rossion & Gauthier, 2002). This classical observation contributes to the hypothesis of a face-specific module in visual processing.

Voice is the carrier of speech but it is also an “auditory face” rich in information on the speaker’s identity, gender, age, familiarity, and affective state. Also, as with faces, voice processing has been linked to selective regions of the auditory cortex (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000), to specific electrophysiological responses (Levy, Granot, & Bentin, 2001) and to a specific recognition deficit in brain-lesioned patients named phonagnosia (Van Lancker, Kreiman, & Cumming, 1989). The present three experiments explored the possibility of a “voice inversion effect” which could support the hypothesis of a voice-specific module. Three types of voice inversion were evaluated: two frequency inversion (spectrogram rotations centered around 2500 and 4000 Hz), and a temporal inversion (e.g., backwards speech). Experiment 1 assessed the effect of voice inversion in a gender identification task. Experiments 2 and 3 compared the effects of inversion on subjects’ performance at two similar auditory discrimination tasks using voices (Experiment 2) and musical instruments (Experiment 3).

2. Material and methods

2.1. Subjects

Twenty subjects (10 women, 10 men) aged between 21 and 23 years old participated in Experiment 1 while 20
different subjects (10 women, 10 men) aged between 20 and 50 years old participated in both Experiments 2 and 3. All subjects were native Canadian–French speakers and none had auditory or neuropsychological problems. They all gave informed written consent.

2.2. Tasks and stimuli

Experiment 1 consisted of a gender identification task on syllables pronounced by 90 speakers (16 boys, 16 girls, 29 men and 29 women). Experiment 2 consisted of a “same/different” speaker discrimination task on pairs of syllables pronounced by 8 men and 8 women. Voice stimuli were excerpts chosen from the recordings of Hillenbrand, Getty, Clark, and Wheeler (1995) and consisted of two syllables pronounced by each speaker: the American vowels “A” and “I” in a /hVd/ context. In Experiment 3, a control condition, subjects performed a “same/different” instrument discrimination task (similar to that of Experiment 2) on pairs of melodies played by 16 instruments (8 strings, 8 wind). Instrument stimuli were synthesized using the MIDI synthesizer of the PC sound card, and two 3-note melodies were obtained from all instruments: C–D–G and C–F–G.

In all three experiments, stimuli were presented in four conditions: (1) no inversion; (2) temporal inversion; (3) frequency inversion centered around 4000 Hz; and (4) frequency inversion centered around 2500 Hz. Pairs of syllables or melodies used in Experiments 2 and 3 were always composed of two sounds from the same category (e.g., two male voices or two string instruments) in the same condition of inversion; the two syllables or melodies of each pair were always different. All stimuli were digital (mono, 16-bit) with sample rate of 10 kHz in conditions “no inversion”, “temporal inversion” and “frequency inversion centered around 2500 Hz,” and 16 kHz in the condition “frequency inversion centered around 4000 Hz.” Original stimuli were inverted in frequency using Matlab 6.1 (Mathworks) and inverted in time using Cool Edit Pro (Syntrellium Software Corporation).

2.3. Procedure

In all experiments, stimuli were delivered binaurally using Beyerdynamics DT770 headphones in a soundproof booth. Sound volume was adjusted to individual comfort level. Stimuli were presented in a pseudo-random order using Media Control Functions (Dig-iVox) in blocks of approximately 15 minutes separated by brief breaks (4 blocs in Experiment 1, and 2 in Experiments 2 and 3). Subjects were instructed to click as quickly as possible in the box related to their answer on the screen (“masculine” vs. “feminine” in Experiment 1; “same” vs. “different” in Experiments 2 and 3) using the left button of the computer mouse. Two response modes were counterbalanced between subjects: in Experiment 1, “feminine” in the left box vs. “masculine” in the left box; and in Experiments 2 and 3, “same” [speaker or instrument] in the left box vs. “different” in the left box. Answers and reaction times were recorded with MCF for analysis. Before the beginning of the experiment, subjects performed 10 practice trials with feedback.

3. Results and discussion

Fig. 1 shows percentage of correct answers for the three experiments according to the conditions of inversion. In Experiment 1, mean percent correct ranged from 78.6 ± 5.1% for the normal condition, to 60.4 ± 7.4% for the 2500-Hz inversion condition. A repeated measure ANOVA showed a significant effect of inversion condition on performance (Greenhouse–Geisser Epsilon $F(2.35, 44.66) = 39.67, p < .001$). Planned paired-sample comparisons showed that performance in the normal condition was also higher than in the temporal inversion condition (mean = 74.1 ± 5.7%, $t(19) = 3.41, p < .005$) and than in the frequency inversion centered around 4000 Hz (mean = 66.5 ± 8.1%, $t(19) = 7.34, p < .001$). Thus, although the different inversion conditions degraded performance for the gender identification task, subjects were still performing above chance.

In Experiment 2 (speaker discrimination task), subjects’ performance was lower than in the gender identification task, ranging from a mean of 67.1 ± 8.9% in the temporal inversion condition, to 59.2 ± 7.1% in the 4000-Hz inversion condition, but was also above chance (Fig. 1). A single-factor repeated-measures ANOVA again showed a significant effect of inversion condition on performance ($F(2.09, 39.72) = 9.44, p < .001$). Planned paired-sample $t$ tests revealed no significant difference between the temporal inversion condition and the normal condition (mean = 67 ± 8.8%; $t(19) = -0.12, p = .91$), but higher performance during the normal condition than during the 4000-Hz inversion condition ($t(19) = 2.97, p < .01$) and the 2500-Hz condition (mean = 59.8 ± 7.1%, $t(19) = 3.76, p = .001$). There was no difference between the two types of frequency inversion ($t(19) = -0.39, p = .7$).

In Experiment 3 (instrument discrimination task), mean percent correct ranged from 94.1 ± 5.1% for the normal condition, to 87 ± 6.9% for the 4000-Hz inversion condition. A single-factor repeated-measures ANOVA again showed a significant effect of inversion condition on performance ($F(2.61, 47.68) = 8.546, p < .001$), with performance in the normal condition higher than in the time-inversion (mean = 90.9 ± 9.1%, $t(19) = 2.17, p < .05$), the 2500-Hz inversion (mean = 89.7 ± 6.9%, $t(19) = 4.03, p = .001$) and the 4000-Hz
inversion (mean = 87.0 ± 6.9%, \( t(19) = 6.13, p < .001 \)) conditions.

Thus, all three experiments indicate a significant effect of sound inversion on the subjects’ performance, with a much stronger effect for frequency inversion than for temporal inversion. This could be explained by the qualitative impression given by the two types of inversion. Indeed, the temporal inversion does not change the timbre of the voice; it only modifies the wording or the melody. On the contrary, the frequency inversion alters greatly the timbre since the sound becomes filled with high frequencies, in a way that might even suggest that the tasks would become impossible to perform. Interestingly however, subjects were still able to identify speaker’s gender or to discriminate speakers or instruments, even in the most disrupting frequency-inversion conditions, since performance was always above chance level (50%).

To be able to conclude on a genuine “voice inversion effect,” one would need to demonstrate a much stronger effect of inversion on the perception of voices than on the perception of non-vocal auditory objects such as musical instruments. Yet, the instrument discrimination task (Experiment 3) yielded higher correct answer rate than the speaker discrimination task (Experiment 2), even for the normal condition (94.1 ± 5.1% vs. 67.0 ± 8.8%, respectively). Such difference in level of difficulty makes it hard to compare the two tasks, since a similar amount of percent correct difference induced by inversion would mean very different things at these different levels of performance. Thus, additional experiments will be necessary before being able to conclude safely on the possible existence of a “voice inversion effect.”

**References**


