

PAPER

## Short-term auditory memory interference: The Deutsch demonstration revisited

Kazuo Ueda\*

*Department of Applied Information and Communication Sciences, Kyushu University,  
4-9-1 Shiobaru, Minamiku, Fukuoka, 815-8540 Japan*

*(Received 11 December 2003, Accepted for publication 23 June 2004)*

**Abstract:** The interference effect of nonspeech and speech in short-term auditory memory was investigated with an experiment paradigm proposed by Deutsch [*Science*, **168**, 1604–1605 (1970)], in which test tones, separated by a 5-s retention interval, were interpolated with six other sounds. In Experiment 1, the test tones were pure tones. The interpolated sounds were pure tones and naturally spoken digits by a female and a male. Nine participants were tested for (1) pure-test-tone pitch recognition, (2) serial recall of the interpolated spoken digits, and (3) both tasks (1) and (2). Pitch recognition errors were significantly increased in task (3) compared to task (1), and the digit recall errors were also significantly increased in task (3). In Experiment 2, the test tones were eight-component harmonic complex tones. The interpolated sounds were eight-component harmonic complex tones, and naturally spoken digits by a male. Twelve participants were tested for the corresponding task conditions as in Experiment 1. Significant increases in the errors of pitch recognition and of digit recall were observed when both tasks were required. These results suggest that speech can interfere with tone pitch in short-term auditory memory, and that pitch salience plays a crucial role in the interference.

**Keywords:** Short-term memory, Interference, Pitch, Spoken numbers, Serial recall

**PACS number:** 43.66.Ba, 43.66.Hg, 43.71.An [DOI: 10.1250/ast.25.457]

### 1. INTRODUCTION

This article is concerned with short-term auditory memory interference for nonspeech and speech. Some researchers (e.g., [1,2]) argue that speech is special and needs a special “speech processor.” If the hypothesis is true, then the special processor might include a specific memory store for speech sounds. That is, there might indeed be a special short-term memory store that only deals with speech, separated from other stores for other categories of sounds.

#### 1.1. Starting Evidence

Work by Deutsch [3] can be interpreted as evidence of the existence of specific pitch stores for tones and spoken numbers. The investigation used the following short-term memory interference paradigm.

The participants were presented with eight successive sounds (Fig. 1). The first one and the last one were the test tones. They were separated by a 5-s retention interval, and were pure tones of 200 ms in duration. The participants had

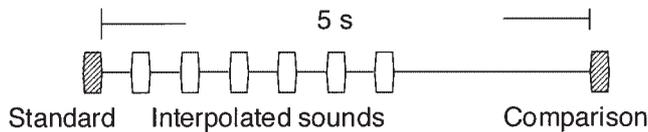
to compare the first tone (the standard tone) and the last (the comparison tone), and to recognize whether they had the same pitch or not. When different pitches were presented, the difference was always a semitone interval, regardless of the direction of pitch change. The standards were randomly selected from an octave range (C $\sharp$ 4–C5; the A4 corresponded to 435 Hz then). In a given trial, the same tone as the standard was presented as a comparison tone with a probability of 50%, and a higher tone and a lower tone were presented with equal probabilities of 25%. Thus, the chance level of performance was 50%. Only the participants who achieved perfect performance in conditions where no sound was presented during the retention interval were selected.

During the retention interval, six other sounds were presented. They were a series of pure tones or spoken numbers. The pure tones were randomly chosen from an octave range, whereas the spoken numbers were randomly chosen from a set of twelve numbers, “one” to “twelve.”

The results were striking: When pitch recognition of the test tones was required and interpolated pure-tones were to be ignored, the percentage of errors was 32.3%. However, the percentage of errors was 2.4% if interpolated

---

\*e-mail: ueda@design.kyushu-u.ac.jp



**Fig. 1** Schematic diagram showing the stimulus order in Deutsch's study [3]. The interpolated sounds were pure tones or spoken numbers. The duration of the pure tones was 200 ms.

numbers were ignored, and was 5.6% if the numbers had to be recalled. The difference between these last two error percentages was not statistically significant. In addition, the number of recall errors were unaffected, no matter whether pitch recognition was required (25.3%) or not (27.4%).

These results can be taken to demonstrate functional specificity in short-term pitch memory: There might be a memory store for pure tones and another for spoken numbers, since interference between spoken numbers and pure tones was virtually none. It should be noted, however, Deutsch herself did not claim such an interpretation was appropriate [3,4].

Pechmann and Mohr [5] argued that Deutsch's results could not be generalized to nonmusicians, since non-musical participants performed the pitch recognition task badly if they were simultaneously required to perform a verbal task (recognition of rhyme in the last two items in a series of interpolated speech sounds). However, their results were consistent with Deutsch's when they tested another group of participants, i.e., participants who reported having some musical experience, but not absolute pitch.

### 1.2. The Importance of Pitch Distance

Pitch distance between test tones and interpolated sounds is an important factor that can seriously affect pitch recognition performance [4,6–8]. Semal and Demany [9,10] conducted experiments that used pure and complex tones, and found that pitch distance between test and interpolated tones mainly determined pitch recognition performance. They also varied sound pressure level, amplitude envelope shape, and spectral composition of interpolated tones, but these factors did not influence performance very much, unless, according to their speculation, the pitch salience of the interpolated tones was affected.

To examine the effect of pitch distance further, another investigation that used harmonic complex tones and spoken French numbers (“sept,” “neuf,” “dix,” and “quinze”) was conducted [11,12]. To define and control pitch of speech, the experimenters used steady tones as guides to be matched by the speaker. The resulting speech samples had relatively flat pitch contours. Fine-tuning of pitch of speech was realized by varying sampling frequency [11,12], and

with a speech analysis-synthesis technique [12]. In this way, they produced a set of complex tones and a set of spoken numbers that were matched with regard to pitch. They compared pitch recognition errors among the conditions that consisted of combinations of tones and spoken numbers as test and interpolated sounds, and found that the pitch distances between test and interpolated sounds primarily determine performance, regardless of sound category.

Therefore, these results can be taken as the evidence that pitch is not stored separately for speech and non-speech.

### 1.3. Purpose

The purpose of this investigation is to clarify further the interference effects between tones and spoken numbers in short-term memory. To this aim, the following five points give the background knowledge worth considering: (1) pitch distances between test and interpolated sounds, (2) pitch salience of speech and nonspeech sounds, (3) use of natural speech and a number recall task, (4) the way of selecting interpolated pure tones, and (5) rejection of absolute pitch possessors with a formal test.

First, to make a fair comparison of interference effects on pitch recognition performance between sound categories, care should be taken over pitch distances between test and interpolated sounds. Since the effect of pitch of speech was not studied by Deutsch [3] and Pechmann and Mohr [5], the confounding effect of pitch could have contaminated their results. Specifically, there was no description concerning pitch of speech in Deutsch [3], therefore one cannot tell even whether the voice used was female or male. There should be generally a large pitch difference between a female voice and a male voice, and the difference is potentially influential on pitch memory interference. Thus, it is very important to check the reproducibility and generality of Deutsch's results with an experiment using both female and male voices.

Second, differences in the pitch salience of speech and nonspeech sounds might have affected the results of Deutsch [3] and Pechmann and Mohr [5]. Pitch of speech generally has large fluctuations. It is quite possible that such fluctuations of pitch of speech cause weaker interference in memory. It is also possible that steady nonspeech sounds may change their pitch salience according to their (fundamental) frequencies.

Third, the experiment that used spoken numbers with a flat pitch contour [11,12] might have caused another problem: Those spoken numbers may not sound speech-like. Since a number recall task was not included in these experiments, it is not obvious (in the sense that there is no behavioral evidence) whether the participants perceived the spoken numbers as speech. Thus, it is necessary to conduct

experiments in which naturally spoken numbers are used and a number recall task is assigned to participants. Such experiments can be regarded as replicating the essence of Deutsch's results.

Fourth, Deutsch [3] selected interpolated pure tones randomly from an octave range, but she excluded tones with the same pitch as test tones. This is because, if an interpolated tone is the same as a test tone, it may enhance test-tone pitch memory [6–8,13]. However, if one uses natural speech sounds as interpolated sounds, and uses test tones that are close to them in pitch, it is almost impossible to avoid pitch overlap between test tones and the interpolated speech sounds, since pitch of natural speech varies continuously over a wide range. Thus, it is also necessary to assess the effect of pitch overlap between test and interpolated tones in order to clarify the difference between tones and speech.

Fifth, previous investigators (including Pechman and Mohr [5]) have not formally checked whether their participants possessed absolute pitch. Absolute pitch possessors are considered to be able to name pitch immediately after they hear a tonal sound [14,15]. Using this ability means that the task is no longer an assessment of pitch-speech interaction, but rather an assessment of linguistic interaction. Thus, absolute pitch possessors should be excluded as participants.

This paper contains two experiments: Experiment 1 is to check the reproducibility and generality of Deutsch's results with pure tones, a female voice, and a male voice; Experiment 2 is an improved version of Experiment 1, using harmonic complex tones of equal loudness and a male voice.

## 2. EXPERIMENT I: PURE TONES VERSUS NUMBERS

### 2.1. Stimuli

Three kinds of stimuli were used in this experiment: (1) pure tones, (2) Japanese numbers uttered by a female speaker, and (3) Japanese numbers uttered by a male speaker.

Pure tones were digitally generated on a workstation (Silicon Graphics, Indy, R4600PC CPU) using 16-bit amplitude quantization and a 12-kHz sampling rate. Their duration was 200 ms including 10 ms rise and fall times. The sound pressure level of the headphones' (STAX, SR  $\Delta$  Professional) output was 75 dB at 1 kHz, and was measured with an IEC coupler (Brüel & Kjaer, Type 4153) and a precision sound level meter (Brüel & Kjaer, Type 2231) in a soundproof room. Pure tones of other frequencies were generated with exactly the same peak amplitude value on the computer.

Spoken Japanese numbers were taken from the ATR Speech Database recorded with 16-bit amplitude quantiza-

tion and a 12-kHz sampling rate. They were uttered by professional speakers, and were recorded in a broadcasting studio [16]. They were edited on the computer to eliminate unnecessary silence before and after the speech sounds.

The stimuli were recorded on DAT (Sony, DTC-95ES), and presented diotically to the participants through headphones. The real-ear frequency response of the headphones measured by loudness comparison showed that the response curve was almost flat from 100 Hz to 20 kHz [17].

#### 2.1.1. Speech loudness equalization

A loudness matching procedure was run to equalize speech loudness. The method of limits was used with a standard pure tone of 1 kHz, 75 dB SPL, and 200 ms duration. Each speech sample was paired with the standard, with a 300-ms inter-stimulus-interval (ISI). The inter-trial-interval (ITI) was 2 s. The presentation order of speech and the standard was inverted after three pairs of ascending and descending series. Thus, for each speech sample, 12 series of measurements were repeated with two participants in a soundproof room. The resulting points of subjective equality (PSEs) of loudness were consistent between the participants: The correlation coefficients of the PSEs were 0.72 for the female speech samples, and 0.97 for the male. The amplitude of each speech sound was adjusted on the computer according to the PSE values averaged over the participants.

#### 2.1.2. Measuring the nominal pitch of the speech sounds

To control pitch distance between test tones and speech sounds, it was necessary to assess pitch of speech. For this purpose, a paired comparison experiment was performed with one participant. Pure tones were paired with each speech sample. The pure tones were spaced every two whole-tones interval. They ranged from C $\sharp$ 3 to C $\sharp$ 5 (138.59–554.37 Hz) for the female speech and from C2 to C4 (130.81–523.25 Hz) for the male speech. The participant had to judge whether a pure-tone pitch was included in a pitch contour of speech, and if it was not, she had to decide which stimulus sounded higher pitched. Measurement was repeated for 20 trials for each combination of the stimulus. The participant's responses were numerically quantified as one point for each "higher" response, and 0.5 point for each "included" response since this implied that pitch of the two stimuli was the same at least at some instances.

A nominal pitch of speech for each speech sample was estimated from the results. The estimated values were averaged for each speaker, and then categorized on a semi-tone scale. The "average" pitch was D $\sharp$ 4 (311.13 Hz) for the female speaker, and was A2 (110.00 Hz) for the male speaker.

### 2.2. Conditions

The framework of the stimulus sequence is indicated in

Fig. 1. Three independent variables were employed: (1) frequency ranges of the standard tones, (2) kinds of interpolated sounds, and (3) tasks for the participants.

- (1) *Frequency ranges of the standard tones.*
  - a. From C $\sharp$ 4 to C5, i.e., 277.18–523.25 Hz. The range is the same regarding pitch names as in Deutsch [3], although the frequencies of the tones were slightly shifted up, due to the shift of the standard (A4) frequency.
  - b. From A3 to G $\sharp$ 4, i.e., 220.00–415.30 Hz. The range is centered on the average pitch of female speech.
  - c. From D $\sharp$ 2 to D3, i.e., 77.78–146.83 Hz. The range is centered on the average pitch of male speech.
- (2) *Kinds of interpolated sounds.*
  - a. No interpolation.
  - b. Pure tones I. This condition was to mimic Deutsch's selection of the interpolated tones. Six interpolated pure tones were randomly selected for each trial. The selections were made from a range the endpoints of which exceeded those of the corresponding standard-tone range by a semitone. In a given trial, however, any tones matching to the standard and its semi-tone above and below were excluded.
  - c. Pure tones II. Six pure tones were completely randomly selected for each trial from the same ranges as in pure tones I.
  - d. Numbers uttered by a female speaker. For a given trial, six numbers were randomly selected out of 10 digits (Japanese "zero" to "nine") for each trial.
  - e. Numbers uttered by a male speaker, selected in the same manner as the above condition.
- (3) *Tasks for the participants.*
  - a. Pitch-recognition task. Pitch recognition of test tones was required but interpolated sounds could be ignored.
  - b. Number-recall task. Number recall for interpolated spoken numbers was required but with no pitch recognition of test tones.
  - c. Dual task. Both pitch recognition and number recall were required.

The independent variables were cross-combined to form the experimental conditions. However, a completely factorial combination was impossible, since one cannot require "number recall" for a no-interpolation condition and for a pure-tone interpolated condition, for example. It should be also noted that, in the pure-tone interpolated conditions, the test tones and the interpolated tones were always selected from the same range of frequencies, whereas in the number interpolated conditions, the test

tones alone changed their frequency ranges.

### 2.3. Participants

#### 2.3.1. Pitch-recognition test and selection of participants

A pitch-recognition test was undertaken by all participants. This test was meant to select participants who could recognize pitch accurately.

A standard tone was followed by a comparison tone with a 5-s retention interval. The standards were 36 pure tones on a semi-tone scale (C $\sharp$ 2–C5, 69.30–523.25 Hz), each of them being presented once in a random order. This yielded 36 trials. The standard and comparison tone were the same in half of the trials, whereas a one semi-tone higher or lower comparison tone was presented with equal probability in another half of the trials.

Thirty-seven participants who reported normal hearing participated in the test. They had to judge whether two stimuli, a standard tone and a comparison tone, had the same pitch or not. They had to write down their responses during a 10-s ITI. They were informed that, if there was a pitch difference, it was always one semi-tone interval.

The criteria to select the participants were as follows: (1) there should be no error in the range of C $\sharp$ 4–C5 (this is the same criterion as in Deutsch [3]); and (2) errors should be less than chance under C4. The reason for introducing the second criterion is as follows. If the selection had been made solely on the first criterion, 25 participants would have passed. However, 18 of these potential participants made more than one error in the range of C $\sharp$ 3–C4, and the number of persons who made more than one error went up to 24 in the range of C $\sharp$ 2–C3. Obviously, it was too strict to apply the same "no error" criterion to the lower ranges.

Twenty participants met the criteria. Among them, nine participants participated in the whole experiment. All of them were undergraduate students at Kyoto Prefectural University.

#### 2.3.2. Absolute pitch test

This test was to exclude participants having absolute pitch ability. Pure tones on a semi-tone scale (C $\sharp$ 4–C5, 277.18–523.25 Hz) were randomly presented with an ITI of 4 s. This yielded 12 trials. Participants were asked to mark a pitch name on a response sheet during an ITI. If they could not give a name, they were asked to mark the "unknown" column. The ITI should have allowed ample time to respond for any absolute pitch possessor, because an absolute pitch possessor can categorize pitch very quickly. It takes only 1.629 s on average for them to select and press a key, even when the stimulus and the response categories were microtonal, i.e., 20-cent interval [14]. It was assumed that an absolute pitch possessor would get a perfect performance for this test, and that was the criterion to reject a participant.

No participant achieved perfect performance, however.

The best participant's percentage of correct responses was 33.3%, the worst was 0.0%, and the average was 17.5%. These performances were comparable to those of the non-absolute-pitch participants described in Miyazaki [14]. Thus, no participant was rejected.

## 2.4. Procedure

Each condition contained 12 trials for each participant. The stimulus order and retention time in a trial are indicated in Fig. 1. A standard tone and interpolated sounds started regularly with 500-ms inter-onset-intervals. The ITI was 10 s. The participants responded during the ITIs. The experiment was run in a quiet room, and was divided into four blocks according to the kinds of interpolated sounds. The no-interpolation condition with C#4–C5 standards was not repeated because it had been already done in Sect. 2.3.1. Thus, 216 trials were run for each participant. Order of the blocks and order of the other conditions within a block were randomized across participants.

In the conditions where only pitch recognition was required, the participants were instructed to ignore interpolated sounds, to answer whether a standard tone and a comparison tone had the same pitch, and not to hum or sing to retain the standard tone pitch.

In the conditions where the participants were required to perform only number recall, they were instructed to always mark the "same pitch" column, and then to write down the digits in the order they heard them. They had to wait for the comparison tone to finish before beginning their response. The participants were informed that, in a given trial, the comparison tone would always be the same as the standard tone.

In the dual-task conditions, the participants were required to respond to pitch before attempting to recall numbers.

## 2.5. Results

The percentages of pitch recognition errors for two kinds of interpolated sounds, i.e., pure tones and spoken numbers, are separately shown in Figs. 2(a) and (b). The percentage errors for the control condition, i.e., no-interpolation condition, are shown in both figures. A correct response for the number recall task, both in the single- and in the dual-task conditions, was to name all six interpolated numbers in the correct order. Otherwise the response was regarded as wrong. Figure 2(c) shows the number recall error percentage. The data by Deutsch [3] are also shown in these figures.

The error rates for each participant were converted into angles with an arcsine transformation [18]. To perform analysis of variance (ANOVA), a GLM (general linear model) procedure in the SAS package was applied. The participants were regarded as a block factor. For the data

summarized in Fig. 2(a), the main effects of frequency ranges of standard tones (R), kinds of interpolated sounds (I), participants (P), and the interaction effect of R × I were significant:  $F(2, 64) = 19.05$ ,  $p = 0.0001$ ;  $F(2, 64) = 41.44$ ,  $p = 0.0001$ ;  $F(8, 64) = 5.48$ ,  $p = 0.0001$ ; and  $F(4, 64) = 5.80$ ,  $p = 0.0005$ , respectively. For the data summarized in Fig. 2(b), the main effects of R, task for the participants (T), and P were significant:  $F(2, 80) = 82.56$ ,  $p = 0.0001$ ;  $F(1, 80) = 17.68$ ,  $p = 0.0001$ ; and  $F(8, 80) = 3.53$ ,  $p = 0.0015$ , respectively, whereas the main effect of I and the interaction effects of R × I, R × T, and I × T were not significant:  $F(2, 80) = 0.47$ ,  $p = 0.6246$ ;  $F(2, 80) = 0.34$ ,  $p = 0.7110$ ;  $F(2, 80) = 1.58$ ,  $p = 0.2121$ ; and  $F(1, 80) = 0.22$ ,  $p = 0.6387$ , respectively.

In the no-interpolation condition, the percentage of pitch recognition errors was high for low frequency test tones, but small or absent for high frequency test tones; whereas in the pure-tone interpolated conditions, the error percentages were consistently high over the whole frequency range of test tones (Fig. 2(a)). The significant interaction effect of R × I was mainly attributed to the difference between these two kinds of conditions, because the differences between pure tone types I and II were negligible.

The error percentages were small when the interpolated sounds were spoken numbers and the participants were required to perform pitch recognition of the test tones but were instructed to ignore the numbers (Fig. 2(b)). They were in essence not different from those in the no-interpolation conditions. However, when the participants were assigned both pitch recognition and number recall tasks, the error percentages increased.

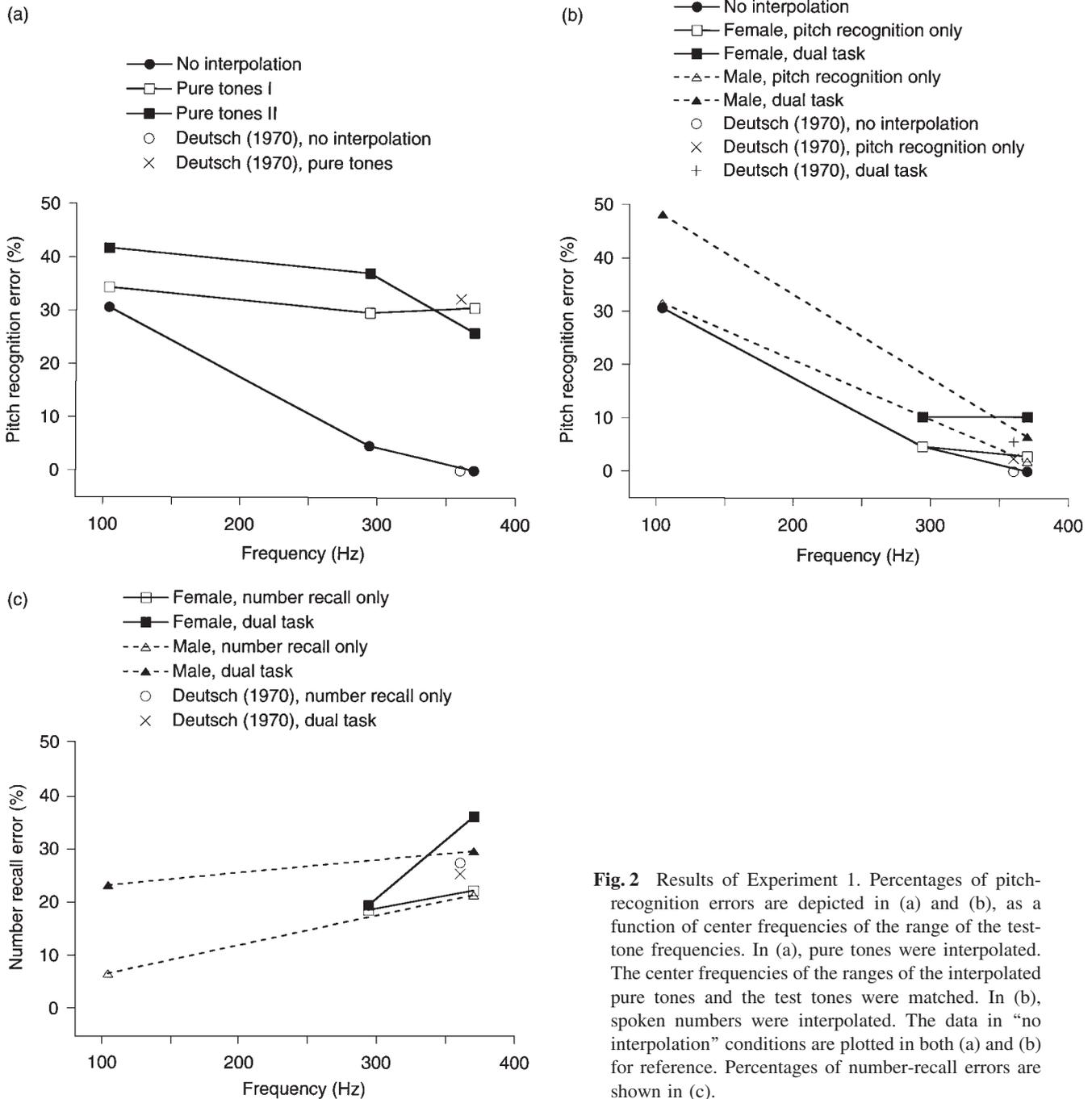
The dual-task effect was also evident in number-recall errors. The error percentages decreased for low frequency test tones (Fig. 2(c)). The main effects of R, T, and P were significant:  $F(2, 56) = 7.80$ ,  $p = 0.0010$ ;  $F(1, 56) = 15.77$ ,  $p = 0.0002$ ; and  $F(8, 56) = 12.51$ ,  $p = 0.0001$ , respectively. The main effects of I and the interaction effects of R × T and I × T were not significant:  $F(1, 56) = 0.87$ ,  $p = 0.3536$ ;  $F(2, 56) = 2.06$ ,  $p = 0.1373$ ; and  $F(1, 56) = 0.40$ ,  $p = 0.5285$ , respectively.

The mean error percentages obtained by Deutsch [3], especially the results of pitch recognition, came close to the present results (Figs. 2(a)–(c)).

## 2.6. Discussion

### 2.6.1. Interference between pure tones and spoken numbers

Significant increases in errors were observed both in pure-tone pitch recognition and in number recall, when the participants were required to perform both the pure-tone-pitch-recognition task and the number-recall task. This fact contradicts the assertion that, on the grounds of the data by



**Fig. 2** Results of Experiment 1. Percentages of pitch-recognition errors are depicted in (a) and (b), as a function of center frequencies of the range of the test-tone frequencies. In (a), pure tones were interpolated. The center frequencies of the ranges of the interpolated pure tones and the test tones were matched. In (b), spoken numbers were interpolated. The data in “no interpolation” conditions are plotted in both (a) and (b) for reference. Percentages of number-recall errors are shown in (c).

Deutsch [3], there exists specific pitch stores for tones and spoken numbers. It is also indicated that the process of interaction is not automatic, because when number recall was not required, spoken numbers did not interfere much with test-tone pitch.

2.6.2. The effect of indistinct pitch on pitch recognition and on number recall

To explain the results in further detail, another point should be taken into account: that is, pure-tone pitch indistinctness at low frequency. Under 1 kHz, the relative differential limen of pure-tone pitch (i.e.,  $Df/f$ ) increases when the frequency decreases. Even for well trained participants, the relative differential limen at 125 Hz is

three times larger than at 1 kHz [19,20]: Pure-tone pitch becomes less distinct at low frequencies.

This indistinct pure-tone pitch may explain why pitch recognition became inaccurate at lower frequencies, even when no interpolated sounds were present. The same line of reasoning may be applied to illustrate the decreases of number-recall error at the lower range of test tones: It is also likely that pitch indistinctness leads to weaker memory for test tones, hence weaker interference with memory for numbers.

2.6.3. The way of selecting interpolated pure tones

Concerning the interpolated pure tones of types I and II, practically no difference was observed in pitch-

recognition-error percentages between the two conditions. Thus, inclusion of the test-tone frequency in the sequence of interpolated tones did not have much effect on the results. However, it is conceivable that a potential difference did not appear because the error percentages at a low frequency are too close to the upper limit (50%).

#### 2.6.4. Comparison with Deutsch's results

The average error percentages of pitch recognition shown in Figs. 2(a) and (b) agree fairly well with Deutsch's results in the corresponding conditions. Her results were replicated in a sense. However, the experiment conditions employed in this investigation were wider than Deutsch's, and the statistical conclusion of the present investigation was inconsistent with hers: There were significant performance differences between the single-task conditions and the dual-task conditions. The task effect was also statistically significant in the number-recall error performance, and is again inconsistent with Deutsch's conclusion.

#### 2.6.5. Musical ability of the participants and rejection of absolute pitch possessors

Our participants made no mistakes in the pitch-recognition test in the range of C#4–C5. Therefore, they can be regarded as musical participants without any absolute-pitch abilities. However, their performance significantly deteriorated when both pitch recognition and number recall were required. This seems to contradict the findings by Pechmann and Mohr [5], who claimed that the pitch-recognition performance by musical participants was not significantly affected even when a verbal task was loaded. One possible explanation may be that, in Pechmann and Mohr [5], change of performance was overshadowed by performance variability of the participants, because that source of variance was not taken into account in their statistical analysis. Since errors are generally fewer in musical participants than in nonmusical participants, this factor can be critical for statistical significance.

### 3. EXPERIMENT II: HARMONIC COMPLEX TONES VERSUS NUMBERS

In Experiment 1, pure tones were used as test tones and interpolated tones. At low frequencies pure-tone pitch might have become less sharply defined. Causing many errors, this indistinctness may have some confounding effects on the results. Furthermore, lower-frequency test tones might suffer less interference from number recall than higher-frequency test tones, in spite of closer pitch distance between the test tones and the speech sounds.

Therefore, in Experiment 2, harmonic complex tones were used instead of pure tones, because their pitches are clearer even at low fundamental frequency. Moreover, the loudness of harmonic complex tones was equalized. These modifications were expected to further clarify the interference effect between nonspeech sounds and speech sounds.

To simplify the experimental conditions, other modifications were also introduced: All the speech sounds were from a male speaker, and the fundamental frequency range of interpolated nonspeech sounds, i.e., harmonic complex tones, was always centered on the frequency corresponding to the average pitch of the speech sounds.

#### 3.1. Stimuli

Eight-component harmonic complex tones were used as test tones and interpolated tones. The amplitude of each component at synthesis was determined as:

$$A_i = 2^{-(i-1)}A_1 \quad (1)$$

where  $A_i$  is the amplitude of  $i$ th component, and  $A_1$  is the amplitude of the first component (the fundamental). They were synthesized on a workstation (Silicon Graphics, Indy, R4600PC CPU) using 16-bit amplitude quantization and a 12 kHz sampling rate. The duration of each tone was 200 ms including 10-ms rise and fall times. The fundamental frequencies ranged from 73.42 to 155.56 and from 261.63 to 554.37 Hz. These ranges correspond to D2–D#3 and C4–C#5, respectively, when A4 is 440 Hz. Twenty-eight tones were generated with a semitone-step interval.

To equalize loudness of the complex tones, the PSE for each tone's loudness was measured with the method of limits. The standard tone was a pure tone of 1 kHz, 75 dB SPL, and 200 ms. The comparison tones were complex tones of which the SPL was changed in a 2-dB step. The ISI was 300 ms, and the ITI was 2 s. Ascending and descending series were alternately presented. Series of both directions were presented four times. In half of the series, the presentation order of standard and comparison tones was reversed. The SPL of the initial comparison tone in a series was randomly chosen from three possibilities ranging over 3 dB, with a 1-dB step. Two participants participated in the measurement. Using average PSEs over the participants, the amplitude of each complex tone was adjusted to yield equal loudness.

The same male speech sounds as in Experiment 1 were used as interpolated sounds.

#### 3.2. Conditions

Three independent variables were employed: (1) fundamental frequency range of the standard tones, (2) kinds of interpolated sounds, and (3) tasks for the participants.

##### (1) Fundamental frequency ranges of the standard tones.

- From C#4 to C5, i.e., 277.18–523.25 Hz. This corresponds to the range used by Deutsch [3].
- From D#2 to D3, i.e., 77.78–146.83 Hz. This range is centered on the average pitch of the male voice.

##### (2) Kinds of interpolated sounds.

- a. No interpolation.
- b. Harmonic complex tones I. Six interpolated tones for each trial were randomly selected from the range of D2–D $\sharp$ 3, 73.42–155.56 Hz, and were arranged to form a series. However, tones within a range of plus or minus one semitone from the trial's standard and tones with the same tone-chromas as the trial's standard, were excluded.
- c. Harmonic complex tones II. The selection was completely random from the same range as for the harmonic complex tones I.
- d. Numbers uttered by a male speaker. Six numbers for each trial were randomly chosen from ten speech samples.

(3) *Tasks for the participants.* They were the same as in Experiment 1.

- a. Pitch-recognition task.
- b. Number-recall task.
- c. Dual task.

The independent variables were cross-combined to form the experimental conditions. The no-interpolation condition was included in the process of participant selection.

### 3.3. Participants

Thirteen participants, who were undergraduate students at Kyoto Prefectural University, participated in the experiment. All of them had normal hearing ability according to screening tests with an audiometer (Dana, DA-301). Pitch-recognition tests of 24 trials, i.e., no-interpolation conditions, were run for all the participants. Twelve participants who committed no error within the range of the C $\sharp$ 4–C5 standard tones continued to the later experimental sessions. This criterion for selecting participants was the same as in Deutsch [3].

The method of absolute-pitch test was the same as in Experiment 1, except that this time the stimuli were harmonic complex tones. It was found that all the participants were non-absolute-pitch-possessors [14]: The best participant's percentage of correct responses was 33.3%, the worst was 0.0%, and the average was 7.8%.

### 3.4. Procedure

Stimuli were recorded on DAT (Sony, DTC-95ES), and were presented diotically through headphones (STAX, SR  $\Lambda$  Professional) to a participant in a sound proof booth (Dana, DA-0020). The background noise level (A weighted) in the booth was measured with a sound level meter (ONO SOKKI, LA-210), and was about 40 dB.

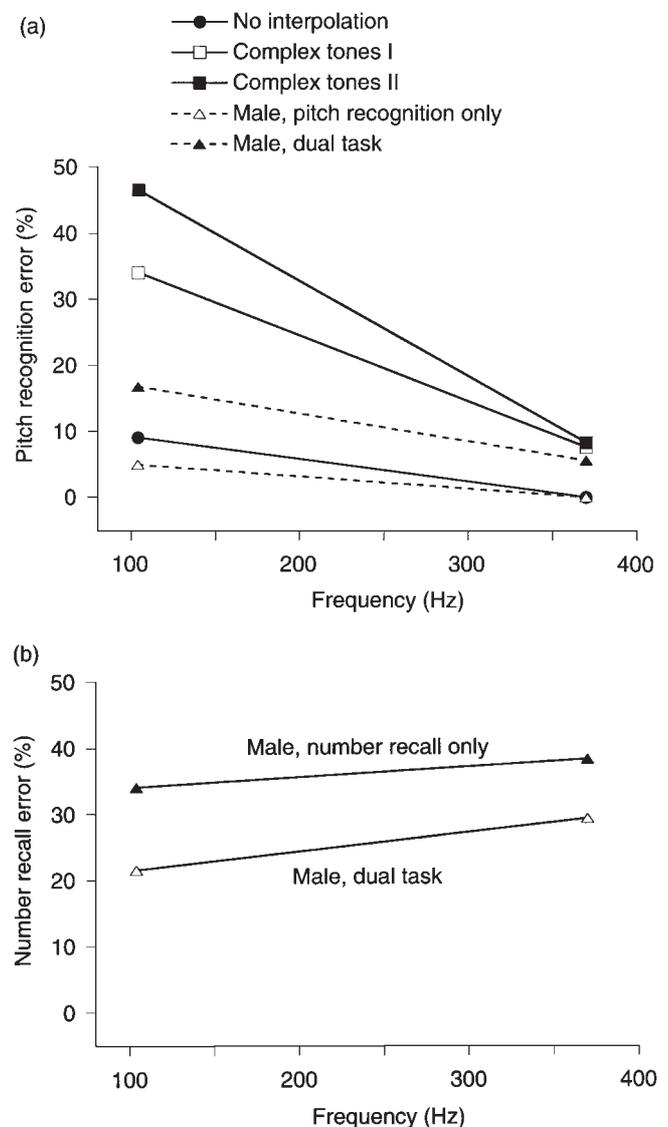
The framework of stimulus order and timing was the same as in Experiment 1. Each condition contained 12 trials for each participant. The experiment was divided into 4 blocks according to the kinds of interpolated sounds. The

order of the blocks and the order of the other conditions within a block were randomized across participants.

### 3.5. Results

Percentages of pitch-recognition errors and number-recall errors are shown in Fig. 3(a) and (b), respectively.

ANOVA was performed on the arcsine transformed error rates, in the same way as for Experiment 1. As for the pitch recognition error rates, the main effect of the frequency range of test tone fundamentals (R),  $F(1, 99) = 110.56$ ,  $p = 0.0001$ ; kinds of interpolated sounds (I),  $F(3, 99) = 40.17$ ,  $p = 0.0001$ ; tasks for the participants (T),  $F(1, 99) = 12.69$ ,  $p = 0.0006$ ; participants (P),  $F(11, 99) = 3.39$ ,  $p = 0.0005$ ; and the interaction effect of R  $\times$  I,  $F(3, 99) = 12.38$ ,  $p = 0.0001$ ; were significant. The interaction effect of R  $\times$  T,  $F(1, 99) = 1.07$ ,



**Fig. 3** Results of Experiment 2. Percentages of pitch-recognition errors are shown in (a), and percentages of number-recall errors are shown in (b).

$p = 0.3029$ , was not significant.

The transformed pitch-recognition error rates were split into two parts according to the two test tone frequency ranges. Multiple comparison (Tukey) at a significance level of 5% for each part of the data revealed that, for the lower range of test tones, there were significant differences in error rate between the conditions where harmonic complex tones I or II interpolated, and the other conditions. The difference between the condition where spoken numbers interpolated but only pitch recognition was required, and the dual-task condition, was also significant. The difference between the conditions of harmonic complex tones I and II was insignificant. For the higher range of test tones, differences between the conditions of harmonic complex tones I or II interpolated and of no interpolation, and also between the conditions of harmonic complex tones interpolated and of spoken numbers interpolated with only pitch recognition required, were significant. Again, no significant difference was observed between harmonic complex tones I and II.

As for the number-recall error rates, ANOVA showed that the main effects of T and P were significant:  $F(1, 33) = 8.45$ ,  $p = 0.0065$ ; and  $F(11, 33) = 11.59$ ,  $p = 0.0001$ , respectively. The main effect of R and the interaction effect of  $R \times T$  were not significant:  $F(1, 33) = 2.41$ ,  $p = 0.1304$ ; and  $F(1, 33) = 0.26$ ,  $p = 0.6145$ , respectively.

When the fundamental frequency of the test tone was in the higher range, pitch-recognition errors were low (Fig. 3(a)). This tendency was most obvious in the complex-tone interpolated conditions. The same tendency was observed for the spoken numbers interpolated conditions and for the no-interpolation conditions, although the size of the effect was much smaller. Regardless of test tone frequency, the harmonic complex tones had a larger interpolation effect on pitch recognition than did the speech sounds.

The task conditions also affected pitch-recognition errors: The error percentages in the single-task conditions were not different from those in the no-interpolation conditions, and were always smaller than those in the dual-task conditions (Fig. 3(a)). The dual-task effect was also observed in the number-recall error (Fig. 3(b)). Nevertheless, number-recall error was not affected by the frequency of the test tone fundamentals.

### 3.6. Discussion

#### 3.6.1. Memory interference between harmonic complex tones and spoken numbers

Pitch memory of harmonic complex tones interferes with the memory of a spoken number series, both pitch-recognition and number-recall performances went down in the dual-task conditions, compared with the single-task

conditions. This is in line with the conclusion of Experiment 1.

#### 3.6.2. Pitch distance between test tones and interpolated tones

In this experiment, the fundamental frequencies of the interpolated harmonic complex tones were always centered on the average pitch of the spoken numbers, and the test tone fundamentals alone changed their center frequencies. Therefore it is possible to assess the effect of pitch distance between the test tones and the interpolated tones on pitch recognition error.

When the interpolated complex tones were present, pitch recognition error percentages were greatly affected by the pitch distance between the test tones and the interpolated tones: Closer distance led to larger interference. This fact further confirmed the importance of the pitch distance for pitch recognition error [4,6–12,21–23].

#### 3.6.3. Pitch recognition error and pitch indistinctness of speech

The interference effect of spoken numbers on pitch recognition of test tones was weaker than the effect of the harmonic complex tones. The pitch-recognition error percentages of the single-task condition did not significantly differ from those in the no-interpolation conditions. For the dual-task conditions, although pitch recognition error percentages increased significantly, they were always less than 20%.

It might be hypothesized that pitch of speech matching or crossing over standard-tone pitch reinforces the memory of the standard-tone pitch (e.g., [6–8,13]), and hence has a weaker interference effect. If this is true, the effect of standard-tone pitch reinforcement should be seen more clearly when the interpolated sound is a harmonic complex tone, because the effect should be stronger when the interpolated pitch is stable. In that case, harmonic complex tones II may interfere less with test-tone pitch than harmonic complex tones I. In fact, the results did not support this hypothesis: There was no significant difference of performance between these two conditions.

Another possible source of weaker interference would be the indistinct pitch of speech sounds. It is difficult to perceive a single, clear, and stable pitch for a spoken word, because of its large pitch fluctuation. Thus pitch memory of speech tends to be vague, and hence memory of test-tone pitch does not suffer strong interference. To check this hypothesis, it is necessary to perform an experiment in which one can change pitch salience of speech by using a sophisticated speech analysis-synthesis technique. This line of investigation has been undertaken in the author's laboratory [24,25].

Yet another possible explanation might be that stream segregation [26,27], in this case, perceptual grouping and separation between speech and nonspeech, prevents inter-

ference. This explanation will be further considered in the General Discussion section.

#### 3.6.4. Number-recall error and harmonic complex tones

The error percentages of number serial recall increased by about 10% compared with the previous experiment using pure tones as the test tones. Nevertheless, the error percentages became independent of the frequency range of the test tone fundamentals. These results can be explained by the more robust memory of test tones. Since pitch of harmonic complex tones is more distinct than pure tones of the same (fundamental) frequency, the memory of the test-tone may have become more robust. The loudness equalization may have also contributed somewhat. Robust memory of the test-tone would have caused a stronger interference effect on the memory of the interpolated spoken number series.

### 4. GENERAL DISCUSSION

The present investigation offered supporting evidence that the short-term memory store is unspecialized for tone pitch and for spoken number series. In other words, both memories have at least one common place of interaction at a certain stage of coding, storing, and retrieving in the memory process. Contrasting the present investigation with Deutsch's experiment [3], the present investigation drew its conclusions based on the results from wider ranges of conditions.

To specify at which stage the interference occurs is beyond the scope of this investigation. However, the following three points should be discussed: (1) perceptual similarity of the stimuli, (2) perceptual organization, and (3) robustness of memory.

Memory of a similar stimulus would interfere more than that of a dissimilar stimulus [28]. This view is supported by the fact that closer pitch distance between the test tones and the interpolated sounds caused larger interference in pitch recognition, and the sound of the same kind, i.e., tones vs. tones, caused larger interference than that of the different kinds, i.e., tones vs. speech. Note that, however, if the similarity is extremely high, it is possible that the memory is reinforced [13]. Also, tones in an octave relationship sound similar and interfere more than tones with a shorter pitch distance [21]. Although such conditions were not included in the present investigation, these facts limit the generality of the conclusion.

Perceptual similarity is closely related to perceptual organization of the stimuli. Perceptual organization with various cues, such as sound source orientation, temporal and timbre grouping, also affect interference [27]. Thus, if test tones and interpolated sounds are segregated effectively, the interference effect decreases. Take an example of the condition where the interpolated sound comprises spoken numbers and the task is just to recognize test-tone

pitch. If complete segregation between the test tones and the interpolated sound is successful with negligible cost for the participants, all they have to do is to retain only one item: the pitch of a standard tone. This makes the condition essentially the same as a no-interpolation condition, and hence performance is the same.

Furthermore, robustness of memory should be taken into account: Indistinct pitch leads to weak memory, and to weak interference on the memory for other sounds. The support for this assertion comes from the following facts. First, pure tones of lower frequency, of which pitch is less distinct, showed weaker interference on speech in a number-recall task than pure tones of higher frequency and distinct pitch. Moreover, harmonic complex tones of equal loudness showed greater interference effects than pure tones. The fact that the interpolated speech sounds showed weak interference on pitch recognition can also be accounted for by their indistinct pitch evoking a weak memory trace.

These three factors, although the first two are sometimes difficult to separate, can affect the interference effect, and it is quite reasonable to conclude that interference between speech and nonspeech was observed in pitch recognition and in serial recall of numbers.

### ACKNOWLEDGMENTS

The author would like to express his gratitude and thanks to Jonathan Goodacre, Laurent Demany, Dylan Jones, Adrian Houtsma, Yoshitaka Nakajima, and two anonymous reviewers who provided many helpful comments on the previous version of this article, and to Yoh'ichi Tohkura for his support. Thanks are due to Wakami Adachi and Tomomi Nagatani for their assistance in running the experiments. This work was supported by Grant-in Aid for Scientific Research No. 06710047 and 08710052, from the Ministry of Education, Science and Culture, with additional support from Nagai Research Found, the Japan Society for the Promotion of Science (Grant No. 14101001), and the governmental "Center of Excellence" (COE) program in Kyushu University entitled "Design of Artificial Environments on the Basis of Human Sensibility." Portions of this article were presented at the Third Joint Meeting of the Acoustical Society of America and the Acoustical Society of Japan, 3 December 1996, in Hawaii.

### REFERENCES

- [1] A. M. Liberman and I. G. Mattingly, "The motor theory of speech perception revised," *Cognition*, **21**, 1–36 (1985).
- [2] J. D. Trout, "The biological basis of speech: What to infer from talking to the animals," *Psychol. Rev.*, **108**, 523–549 (2001).
- [3] D. Deutsch, "Tones and numbers: Specificity of interference in immediate memory," *Science*, **168**, 1604–1605 (1970).

- [4] D. Deutsch, "The organization of short-term memory for a single acoustic attribute," in *Short-term Memory*. D. Deutsch and J. A. Deutsch, Eds. (Academic, New York, 1975), pp. 107–151.
- [5] T. Pechmann and G. Mohr, "Interference in memory for tonal pitch: Implications for a working-memory model," *Mem. Cognit.*, **20**, 314–320 (1992).
- [6] D. Deutsch, "Mapping of interactions in the pitch memory store," *Science*, **175**, 1020–1022 (1972).
- [7] D. Deutsch, "Facilitation by repetition in recognition memory for tonal pitch," *Mem. Cognit.*, **3**, 263–266 (1975).
- [8] D. Deutsch and J. Feroe, "Disinhibition in pitch memory," *Percept. Psychophys.*, **17**, 320–324 (1975).
- [9] C. Semal and L. Demany, "Dissociation of pitch from timbre in auditory short-term memory," *J. Acoust. Soc. Am.*, **89**, 2404–2410 (1991).
- [10] C. Semal and L. Demany, "Further evidence for an autonomous processing of pitch in auditory short-term memory," *J. Acoust. Soc. Am.*, **94**, 1315–1322 (1993).
- [11] K. Ueda, C. Semal and L. Demany, "Tones and numbers: Little specificity of interference in immediate memory," *Trans. Tech. Comm. Psychol. Physiol. Acoust.*, H-93-77 (1993).
- [12] C. Semal, L. Demany, K. Ueda and P. Hallé, "Speech versus nonspeech in pitch memory," *J. Acoust. Soc. Am.*, **100**, 1132–1140 (1996).
- [13] D. Deutsch, "Effect of repetition of standard and comparison tones on recognition memory for pitch," *J. Exp. Psychol.*, **93**, 156–162 (1972).
- [14] K. Miyazaki, "Musical pitch identification by absolute pitch possessors," *Percept. Psychophys.*, **44**, 501–512 (1988).
- [15] W. D. Ward, "Absolute Pitch," in *The Psychology of Music*, 2nd ed., D. Deutsch, Ed. (Academic Press, San Diego, 1999), pp. 265–298.
- [16] Y. Sagisaka and N. Uratani, "ATR spoken language database," *J. Acoust. Soc. Jpn. (J)*, **48**, 878–882 (1992).
- [17] K. Ueda and T. Hirahara, "Frequency response of headphones measured in free field and diffuse field by loudness comparison," *J. Acoust. Soc. Jpn. (E)*, **12**, 131–138 (1991).
- [18] G. W. Snedecor and W. G. Cochran, *Statistical Methods*, 8th ed. (Iowa State University Press, Ames, 1989).
- [19] J. O. Nordmark, "Mechanisms of frequency discrimination," *J. Acoust. Soc. Am.*, **44**, 1533–1540 (1968).
- [20] B. C. J. Moore, "Frequency difference limens for short-duration tones," *J. Acoust. Soc. Am.*, **54**, 610–619 (1973).
- [21] D. Deutsch, "Octave generalization of specific interference effects in memory for tonal pitch," *Percept. Psychophys.*, **13**, 271–275 (1973).
- [22] D. Deutsch, "Interference in memory between tones adjacent in the musical scale," *J. Exp. Psychol.*, **100**, 228–231 (1973).
- [23] D. Deutsch, "Generality of interference by tonal stimuli in recognition memory for pitch," *Quart. J. Exp. Psychol.*, **26**, 229–234 (1974).
- [24] K. Ueda and N. Seo, "Short-term auditory memory interference: the effect of speech pitch salience," *Proc. 16th Int. Congr. Acoust. and 135th Meet. Acoust. Soc. Am.*, pp. 2355–2356 (1998).
- [25] K. Ueda, "Short-term auditory memory interference: Dual task load and streaming," *Proc. 7th West. Pac. Reg. Acoust. Conf.*, pp. 199–202 (2000).
- [26] A. S. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge, Mass., 1990).
- [27] D. M. Jones, W. J. Macken and C. Harries, "Disruption of short-term recognition memory for tones: Streaming or interference?," *Q. J. Exp. Psychol.*, **50A**, 337–357 (1997).
- [28] G. E. Starr and M. A. Pitt, "Interference effects in short-term memory for timbre," *J. Acoust. Soc. Am.*, **102**, 486–494 (1997).