

5aSC12

Abstract in: *Journal of the Acoustical Society of America* 105 (2, part 2):3204 (1997).

Native language determines parsing of nonlinguistic rhythmic stimuli

Kiyomi Kusumoto and Elliott Moreton
Department of Linguistics, University of Massachusetts/Amherst
Contact: elliott@linguist.umass.edu

Abstract

Given a rhythmic sound stimulus, the hearer automatically parses it into recurrent units. Researchers since Woodrow [Arch. Psych. 14, 1-66 (1909)] have consistently found that stimulus elements differing in duration are parsed iambically, while those differing in intensity are parsed trochaically. Hayes [*Metrical Stress Theory* (1995)] observes that no language has quantity-insensitive iambic feet, and suggests that the nonlinguistic parsing tendencies may be at the root of this language universal. However, the relevant nonlinguistic parsing experiments were done with native English speakers. Using comparable nonspeech rhythms, we compared parsing behavior in native speakers of English and in two dialects of Japanese. Our findings suggest that language determines parsing preference, not the other way around.

Introduction

We present a case in which nonspeech perception seemed to be determining speech perception. Our results suggest the reverse -- that speech perception is influencing nonspeech perception.

Parsing preferences for nonspeech rhythms

A long sequence of two alternating sounds (tones, clicks, etc.) is heard as a repeating group of two. Researchers since Woodrow (1909) have consistently found *parsing preferences*:

- If the two elements differ only in **intensity**, they are parsed **trochaically** (LOUD-soft)
- If they differ only in **length**, they are parsed **iambically** (short-LONG)

This was hypothesized to be an intrinsic property of the human auditory system. As such, it seemed related to the following

Language universal (Hayes 1995)

Languages subdivide a word into one- or two-syllable prosodic constituents ("feet"), with either initial prominence (trochees) or final prominence (iambs).

If a language **ignores syllable length** in forming feet, then the feet are all **trochees**

If a language forms **iambs**, then every long syllable is final in its foot.

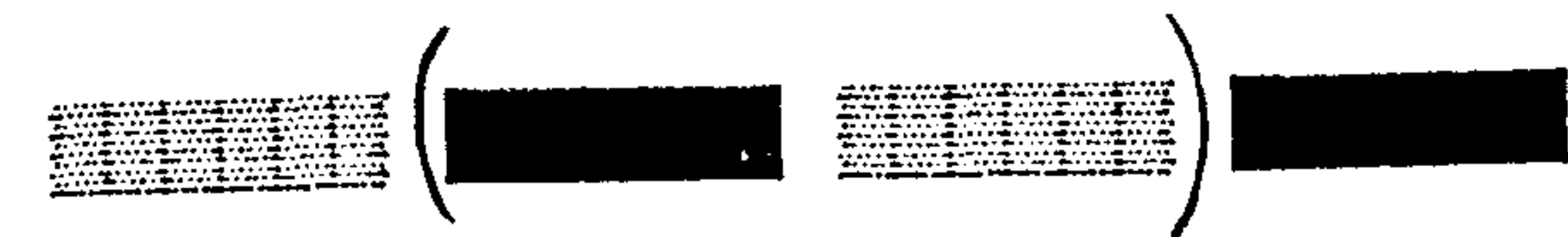
The nonspeech parsing preferences, however, are based on English-speaking subjects. We show that Japanese speakers lack the iambic preference for length contrasts, implying that the parsing preferences are an effect, not a cause, of the rhythmic structure of language.

Background: Rhythmic parsing preferences

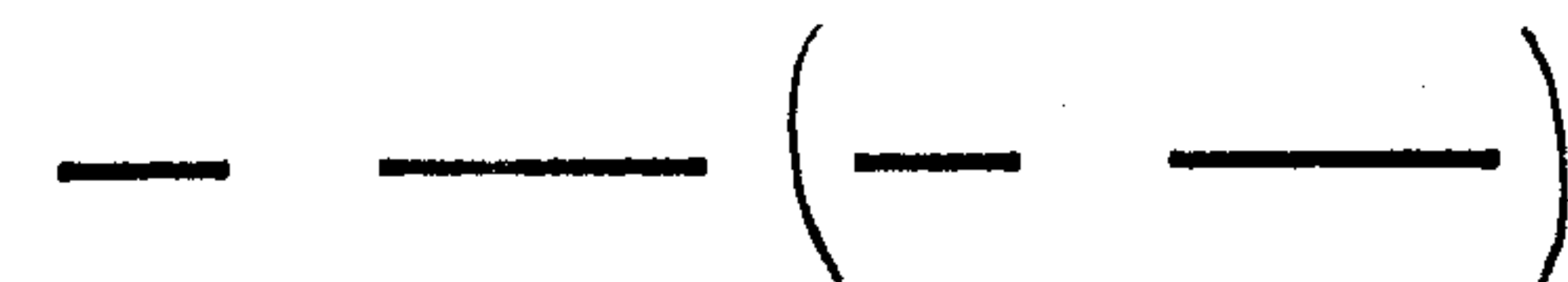
Given a repeated sequence of nonspeech sounds, listeners automatically organize (parse) it into recurrent rhythmic units. There is a strong tendency to distinguish one sound in the unit as salient or prominent over the others.

If the stimulus consists of only two sounds alternating with each other at a moderate rate (between 1/2 and 5 cycles per second), how they are parsed is strongly influenced by how they differ.

- Intensity: Trochaic (LOUD-soft)

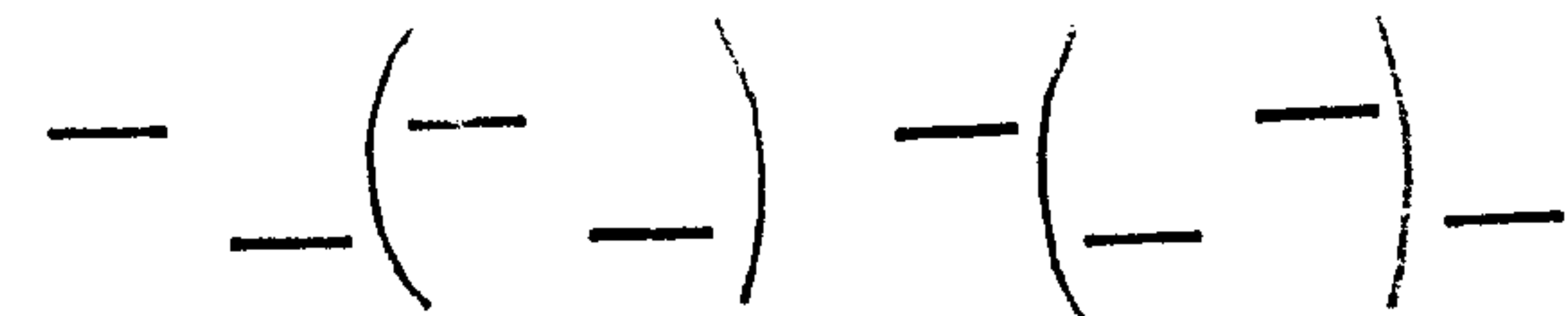


- Duration: Iambic (short-LONG)



If the length ratio is less than about 1.5 to 1, a duration contrast is likely to be perceived and parsed as an intensity contrast (Woodrow 1909, Bell 1977).

- Pitch: Inconsistent (high-low) or (low-high)



These results have been replicated many times (e.g., Woodrow 1909, Bell 1977, Rice 1992), but only with English-speaking subjects. [See Note 1.]

Experiment

We attempted to replicate the nonspeech-parsing-preference results with American English speakers in the United States and speakers of two dialects of Japanese — Kansai and Tokyo — in Japan.

The English speakers performed as expected, parsing intensity contrasts trochaically and duration contrasts iambically.

The Japanese speakers parsed intensity contrasts trochaically, but showed a slight tendency to parse duration contrasts trochaically as well.

Subjects

A total of 25 people participated in the experiment. All reported no hearing or language impairments. Most had had some training in the Western musical tradition, but none was a professional musician. Subjects were native speakers of one of the following three languages:

Kansai Japanese	(K)	10
Tokyo Japanese	(T)	8
American English	(E)	7

Stimuli

Using MacProbe (Aristometrics, Inc.), we synthesized 81 sequences of tones. Each sequence was 14s long — 3.5s fade-in, 7s steady-state, and 3.5s fade-out — and consisted of two alternating sawtooth tones, separated by 15ms silent gaps. Each tone was quickly ramped on and off to eliminate popping.

The three contrasts had each three levels of starkness — the ratio between the magnitudes of the two tones.

INTENSITY ratio of 1.125, 1.50, or 2.00

DURATION ratio of 1.25, 1.75, or 3.00

PITCH ratio of 1.02, 1.20, or 2.00

The stimuli also varied in pitch of the lower tone (100Hz, 250Hz, 500Hz) and duration of the shorter tone (100ms, 150ms, 250ms).

Procedure

The 81 stimuli and 9 practice trials were recorded in random order on high-quality audio tape. Each stimulus was followed by 6s of silence. Stimuli were presented binaurally to Ss on a portable consumer-quality tape player at self-selected listening levels. Ss were encouraged to rewind the tape and listen to each sequence as many times as they wanted before going on to the next stimulus.

Responses were collected in booklets that showed cartoons of the sequences:



One trial per page meant that only the answer blank for the current trial was visible to the subject at any time. Subjects were instructed and practiced in the use of the answer blanks before any experimental trials were run. To counterbalance any influence of the answer blanks themselves on judgments, there were two versions of the book, with complementary arrangements of graphical elements for each trial.

Results

The dependent variable was % trochaic responses -- (LOUD-soft), (LONG-short), or (HIGH-low) as the case might be.

A parsing preference, marked in the tables below with a *, was defined as a mean % trochaic response for which a 95% confidence interval excluded the chance level of 50%.

The findings were:

INTENSITY: All three groups strongly preferred the trochaic parse (LOUD-soft).

DURATION: English speakers, as expected, parsed iambically (short-LONG), but neither group of Japanese speakers showed a reliable preference.

PITCH: English and Kansai speakers had a weak preference for (HIGH-low) trochees; Tokyo speakers were indifferent.

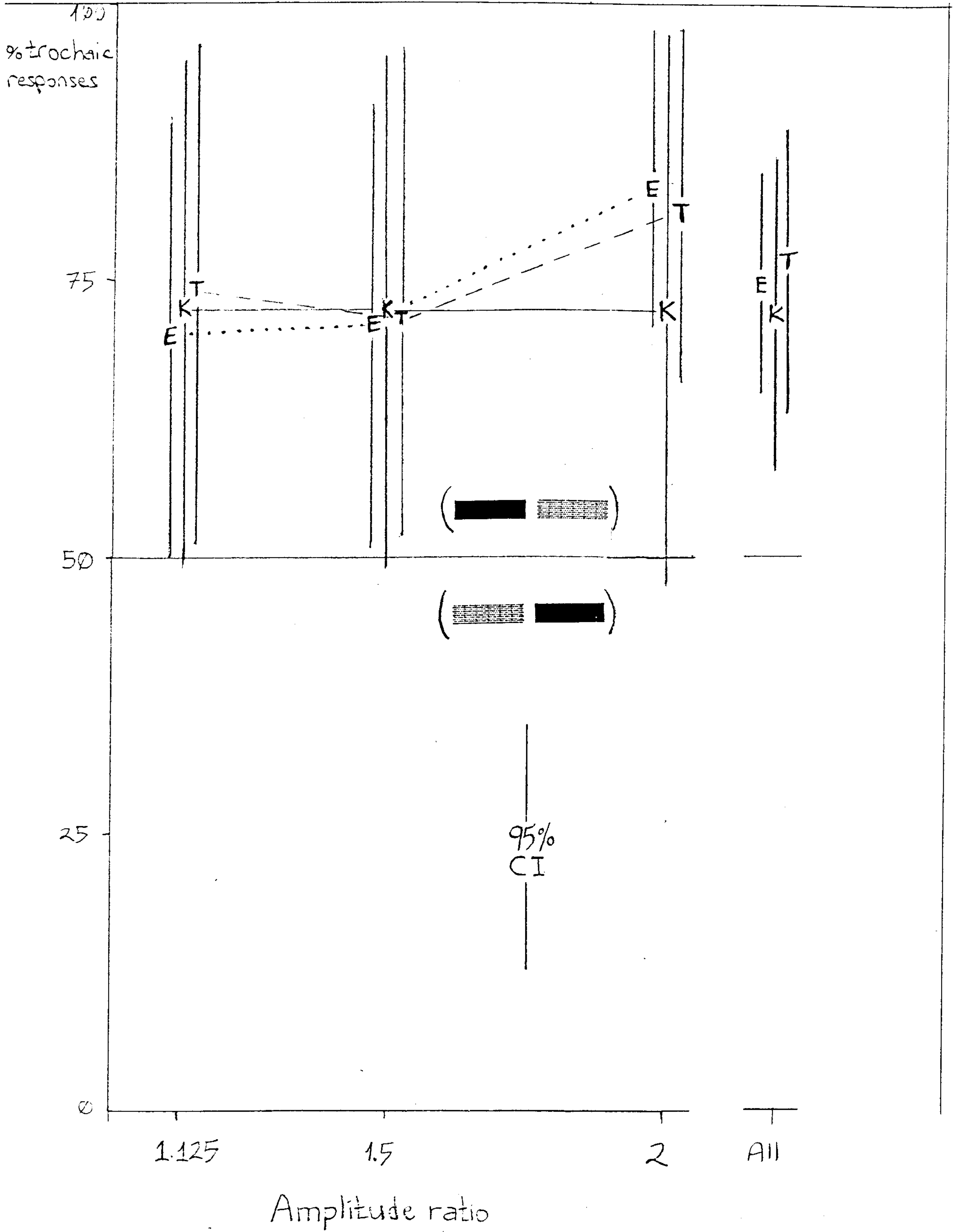
INTENSITY

All three groups showed a trochaic parsing preference, and could not be distinguished from each other.

	Mean	95% CI	
E:	75	[66, 85]	*
K:	72	[58, 86]	*
T:	77	[66, 88]	*

There was a nonsignificant trend towards stronger trochaic preference on starker contrasts.

This finding replicates previous results and establishes the sensitivity of the paradigm.



DURATION

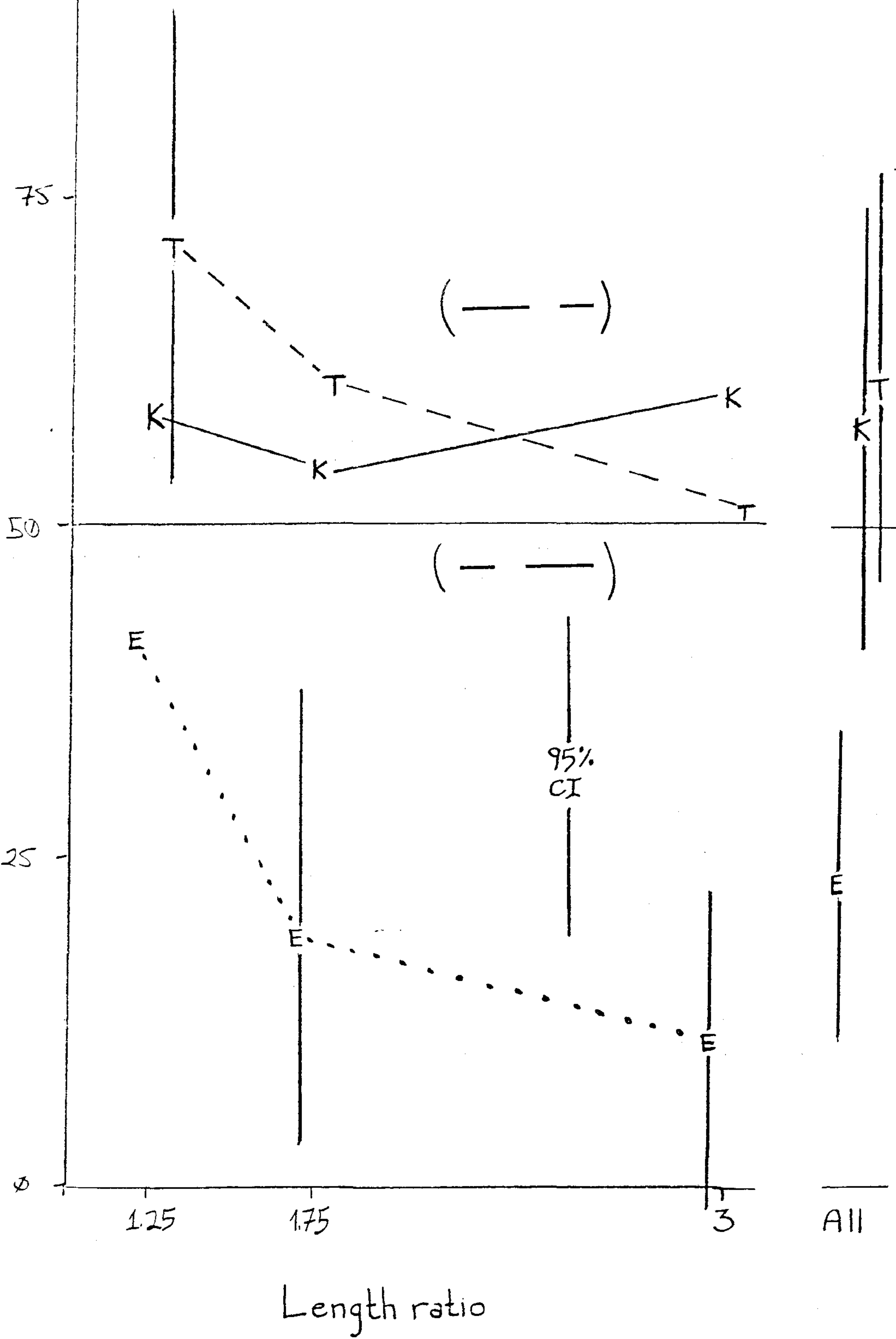
Overall, only English speakers showed a parsing preference — quite a strong one, in the expected iambic direction. Japanese speakers did not.

	Mean	95% CI	
E:	24	[12, 36]	*
K:	57	[40, 74]	
T:	61	[47, 75]	

At the smallest length ratio, 1.25, the Tokyo group actually chose a trochaic parse, but this was probably due to the tendency to hear small length contrasts ($< \sim 1.5$) as amplitude contrasts and parse them accordingly.

The lack of an iambic preference for the Japanese speakers is unexpected, given past results.

100
% trochlear
responses

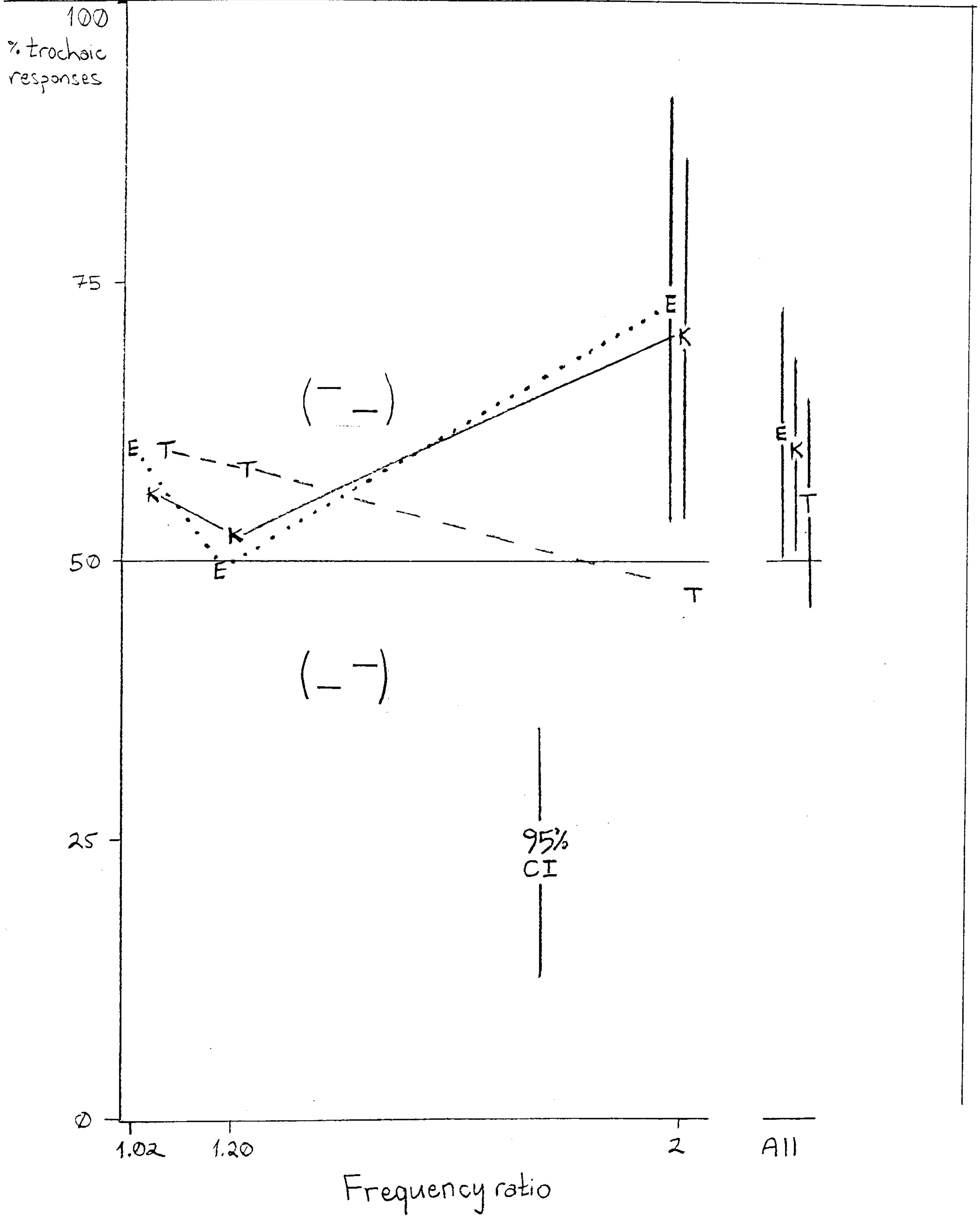


PITCH

The English and Kansai groups slightly favored (high-low) trochees; Tokyo speakers showed no reliable preference.

	Mean	95% CI	
E:	62	[50, 74]	*
K:	60	[51, 69]	*
T:	55	[46, 64]	

This is in keeping with previous reports of fragile pitch parsing preference (Rice 1992).



Conclusions

We have shown that native language can have a strong influence over the perception of non-linguistic stimuli. This effect on rhythm perception is parallel to the effect of language on pitch perception found by Deutsch (1991). Linguistic influences in such dissimilar domains feed the suspicion that more are yet to be found, that the influence of language may pervade auditory perception.

How does language exert this influence on rhythm? The most dramatic effect which we have found, and the one most requiring explanation, is English speakers' strong preference for parsing duration contrasts iambically (short-LONG), a preference which Japanese speakers lack.

We speculate that the difference results from the fact that English prohibits words ending in a short (open) syllable with a full vowel, which Japanese permits. Listeners covertly shadowing the stimulus with reduplicated nonsense words ("*dee-dee dee-dee...*") can form [CVV.CV] and [CV.CVV] words in Japanese, but only the iambic [CV.CVV] in English. Listeners covertly matching the stimulus to a real disyllabic English word will be affected by the almost 5-to-1 preponderance of (short-LONG) over (LONG-short) words in the English vocabulary (CELEX database count).

Note

1. Bell (1977) did test speakers of several very different languages using three-element stimuli, but his duration contrasts (a) used a ratio of 1.4, which makes it likely that they were heard as intensity contrasts, and (b) kept constant the onset-to-onset interval, with the result that the silent gap between tones was shorter after the long element than after the others. People strongly prefer not to put a constituent break in the shortest silent interval (Woodrow 1909, Rice 1992). Not surprisingly, Bell's subjects, whatever their language, parsed the duration contrasts so that the long segment came first. Because of these confounds, this result does not bear on the question of parsing preferences for length contrasts.

Selected references

- Bell, A. (1977) Accent placement and perception of prominence in rhythmic structures. Pp 1-13 in: L. Hyman (ed.), *Studies in Stress and Accent*, Southern California Occasional Papers in Linguistics No. 4. Department of Linguistics, UCLA.
- Deutsch, D. (1991) The tritone paradox: An influence of language on music perception. *Musical Perception* 8(4):335-347.
- Hayes, B. (1995) *Metrical Stress Theory: Principles and case studies*. Chicago: University of Chicago Press.
- Rice, C. (1992) Binariness and ternariness in metrical theory: Parametric extensions. Dissertation, UT-Austin.
- Woodrow, H. (1909) A quantitative study of rhythm. *Archives of Psychology* 14:1-66.

Acknowledgements

Thanks are due to Bruce Hayes, John Kingston , and Lisa Selkirk for their ideas, comments, and advice, and to participants in the experiment for their time. Special thanks to Mary C. Potter for the use of her laboratory in making and analyzing the stimuli.