**Proceedings of the Institute of Acoustics**

# LEXICAL EFFECTS ON PHONEME PERCEPTION IN SPOKEN JAPANESE WORDS

Shigeaki Amano

NTT Basic Research Laboratories,
3-9-11 Midori-cho Musashino, Tokyo 180, Japan
amano@siva.ntt.jp

## 1. INTRODUCTION

The phoneme perception process and the word perception process interact with each other. Phoneme information is used for word perception and likewise word information is used for phoneme perception. Phoneme monitoring studies (e.g., [1], [2], [3], and [4], and [5]) show that phoneme perception is facilitated by lexical information; the reaction times to a phoneme in a word are shorter than in a nonword.

This lexical effects become stronger when the phoneme is located near the end of the word. But when do they begin? Frauenfelder, Segui, and Dijkstra ([4], [5]) adopted the "uniqueness point" as the reference point for word/nonword recognition and used it to investigate when the lexical effects begin. The uniqueness point is defined as the point in a word at which it becomes uniquely identifiable from all other possible words, when word matching is performed by starting at the beginning of the word and then working through to the end using a dictionary. They found that after the uniqueness point, the reaction times for target phonemes in words were much shorter than those in nonwords, indicating a strong lexical facilitation. They also found that before the uniqueness point, the reaction times for target phonemes in words and nonwords were almost the same, showing very weak lexical facilitation. They conducted precise analysis of this weak facilitation and concluded that there were no lexical effects before the uniqueness point.

The reference point for word/nonword recognition can be estimated using not only the uniqueness point, but also using other experimental procedures. One of them is a lexical decision task (e.g., [6], [7], and [8]), which requires subjects to discriminate between words and nonwords as quickly as possible. The reaction times are then measured and the reference point for word/nonword recognition is estimated based on these times.

Another experimental procedure used to estimate the reference point for word/nonword recognition is a gating task ([9], [10], [11]), in which subjects hear only fragments of a word (gate). With each trial, they gradually hear more fragments. They guess the word and rate their confidence for each trial. The reference point for word/nonword recognition is estimated from the gate at which subjects correctly guess a word. There have been a number of studies that used the gating task to explore the word recognition process ([12], [13], [14], [15]).

This study tries to determine the lexical effects on phoneme perception before and after the reference point for word/nonword recognition, which is estimated by not only using the uniqueness point but also by a lexical decision task and a gating task.

## 2. METHOD

### 2.1 Subjects

The subjects were 1 male (29 years old) and 5 female (20-25 years old) native Japanese speakers with normal hearing ability. The male subject is a researcher at the NTT Basic Research Laboratories. The female subjects are also employees at the NTT Laboratories and are highly trained in evaluating the quality and intelligibility of speech.

### 2.2 Materials

The stimuli were all Japanese words and nonwords organized under three conditions: (a) lexical information (word vs. nonword), (b) stimulus length in mora and target phoneme position, and (c) presence of target phoneme /k/.

Lexical information was defined by whether the stimuli were meaningful Japanese words or nonsense words. The Japanese stimulus words were selected from a Japanese dictionary containing about 58,000 words. Familiar words were chosen as often as possible, although word frequency was not controlled exactly. Nonwords were generated by reversing moras in the word stimuli when the word length was more than two moras, or by randomly selecting a mora when the length was two moras. No 1-mora nonwords were used as stimuli because almost all 1-mora Japanese words are meaningful.

A native Japanese female speaker pronounced the stimuli at an approximately constant speaking rate. The accent patterns of the stimuli were "flat type", which means that the first mora pitch was low and the rest were high. The stimuli were recorded at a 16-kHz sampling rate with 16-bit A/D conversion in a soundproof booth with a signal-to-noise ratio of 56.5 dB. The average mora duration was 220 ms. The average power of each stimulus was digitally normalized on a computer after the recording.

The stimulus lengths were 1 to 5 mora. To reduce length ambiguities, all stimulus moras were constructed of a consonant and a vowel (CV), or of only a vowel (V). That is, no stimuli contained a long vowel (*e.g.*, */koRsa/*), an obstruent mora (*e.g.*, */kiQte/*), or a nasal mora (*e.g.*, */kaN/*). These constraints made the mora boundaries equal to the syllable boundaries.

The target phoneme position ranged from the initial mora to the last mora. The number of positions depended on the length of the target-bearing stimuli. For example, there is only one target phoneme position in a 1-mora stimulus, but five in a 5-mora stimulus.

The condition for the presence of the target phoneme /k/ was whether the stimulus had one phoneme /k/ or no phonemes /k/ (i.e., a distractor). The number of distractors was the same as for the target-bearing stimuli.

If these three conditions are orthogonalized, (lexical information) x (target position in 1- to 5-mora stimuli) x (target phoneme existence) = 2 x 15 x 2 = 60 cells are obtained. Because 1-mora nonwords were not used, there were actually 59 cells in total. Each cell for stimulus lengths of two or more contained 15 stimuli. Because there were only five possible targets for 1-mora words (/ka/, /ki/, /ku/, /ke/, and /ko/), they were duplicated six times so that the 1-mora-word target cell had 30 stimuli. The 1-mora-word distractor cell also had 30 stimuli. There were 480 words and 420 nonwords. The total number of stimuli was therefore 900. All stimuli were used for phoneme monitoring and lexical decision task, but only /k/-contained stimuli were used for gating task.

### 2.3 Procedure

A personal computer (NEC, PC-9801VX) with a DMA board (I.O DATA, PIO-9322) controlling a 16-bit D/A converter (PAVEC, MD-8000mkII) was used for presenting the stimuli. The D/A output was filtered by an 8-kHz, 135-dB/octave low-pass filter (NF CIRCUIT DESIGN BLOCK, P-86) and then amplified (SONY, TA-F333ESX). The stimuli were diotically presented to the subject through headphones (STAX, SR-ΛPRO) at 71 dB SPL. The stimulus order was randomized for each subject by the computer.

**2.3.1 Phoneme monitoring.** In the phoneme monitoring task, a pure tone (1000 Hz, 150 ms) was presented first, followed by a 1000-ms foreperiod, and then the stimulus. The subjects were in a listening booth with 24-dB(A) background noise. They were instructed to respond as quickly as possible by pushing one of two keys with their first finger when they heard the target phoneme /k/, and by pushing the other key with their second finger when they perceived that the stimulus did not contain the target phoneme /k/. The subjects were instructed to respond as accurately as possible. The keys were connected to the personal computer by a parallel interface board (I.O DATA, PIO-9022B), and responses were automatically recorded. The reaction times for phoneme /k/ were measured from the burst onset of the phoneme. The interval between successive trials was about 2 s. Each subject did 840 trials after 60 training trials. There was a 3-min break after every 70 trials.

**2.3.2 Lexical decision task.** In the lexical decision task, the subjects were instructed to make a lexical decision about the stimuli as rapidly as possible, and to avoid errors as much as possible. When they heard a word, they pushed the left key with their first finger and when they heard a nonword, they pushed the right key with their second finger. The reaction times for lexical decision were measured from the stimulus onset by the computer. The reference points for word/nonword recognition in the lexical decision task (L.W.P./L.N.P.) were obtained from the lexical decision time for each trial. Other conditions were the same as for phoneme monitoring.

**2.3.3 Gating task.** The length of the initial gate was 165 ms. The successive gates were obtained by adding 55 ms to the previous gate. The end of each gate was shaped by a half-cycle cosine window (20 ms) to prevent an audible click. The last gate was the same as the original stimuli. Subjects heard from the initial gate to the last gate. The interval between successive gates was 3 s. For each gate, subjects wrote down their guess in hiragana (Japanese characters). They also rated their confidence by using a scale of "−10 to 10", in which the plus and minus signs correspond to words and nonwords, respectively. The absolute value corresponds to the confidence level.

An 80% threshold of confidence ([14], [15]) was used in estimating the reference points for word/nonword recognition in the gating task (G.W.P./G.N.P.). The G.W.P. was estimated to be the end time of the gate at which the subjects rate confidence to be more than 8 and after which the subjects do not change their guessed word. The G.N.P. was estimated to be the end time of the gate at which the subjects rate confidence to be less than −8.

**2.3.4 Uniqueness point.** The uniqueness point of a word (U.P.) and of a nonword (N.P.) were calculated using a computer-based mora dictionary containing approximately 60,000 Japanese words. The U.P./N.P. was obtained in units of mora. To compare U.P./N.P. with L.W.P./L.N.P. and G.W.P./G.N.P., which were obtained in units of time, they were converted to time as follows. The times for the beginning and end of the mora of U.P./N.P. were determined from the stimulus waveforms by using a waveform editor on a computer. The mean times of the beginning and end were then assigned to the time of U.P./N.P. When a stimulus did not contain a U.P./N.P., the end time of the stimulus was used as the time of U.P./N.P.

## 3. RESULTS

Erroneous responses are excluded from the following analyses. These are defined as: (a) wrong key pressed in phoneme monitoring or in lexical decision task, (b) reaction time of more than 800 ms or less than 80 ms in phoneme monitoring, (c) reaction time of more than 1700 ms or less than 0 ms in lexical decision task, (d) no correct word guess until the last gate in gating task, and (e) an absolute value of a confidence rating less than eight at the last gate in gating task.

Because the 1-mora stimuli with the phoneme /k/ were presented six times in the phoneme monitoring and lexical decision tasks, the 1-mora data were excluded from the following analyses and only the 2- to 5-mora data were used.

The error rates for target-bearing stimuli in phoneme monitoring, lexical decision task, and gating task were 3.85%, 4.68%, and 2.59%, respectively.

Figure 1 gives the mean reaction times for the phoneme /k/ in phoneme monitoring as a function of the delay between the burst onset of the phoneme and (a) L.W.P./L.N.P., (b) G.W.P./G.N.P., and (c) U.P./N.P. The vertical axis corresponds to $RT(/k/)$ and the horizontal axis corresponds to the delay. The mean reaction times for phoneme /k/ were calculated in 100-ms steps of the delay from the onset of /k/, incorporating all the stimulus lengths and all the subjects. That is, the symbols in Figure 1 represent the mean reaction times in the region of ±50 ms on either side of the point. If there were less than five data inputs contributing to a point, it was assumed to be unreliable and discarded.

The mean reaction times for phoneme /k/ in words and nonwords in Figure 1a differed significantly at −300 ms on the horizontal axis, $F(1,248) = 12.6$, $p < .01$, at −200 ms, $F(1,174) = 16.6$, $p < .01$, at −100 ms, $F(1,88) = 11.1$, $p < .01$, at 0 ms, $F(1,41) = 7.29$, $p < .01$, and at 100 ms, $F(1,19) = 10.5$, $p < .01$.

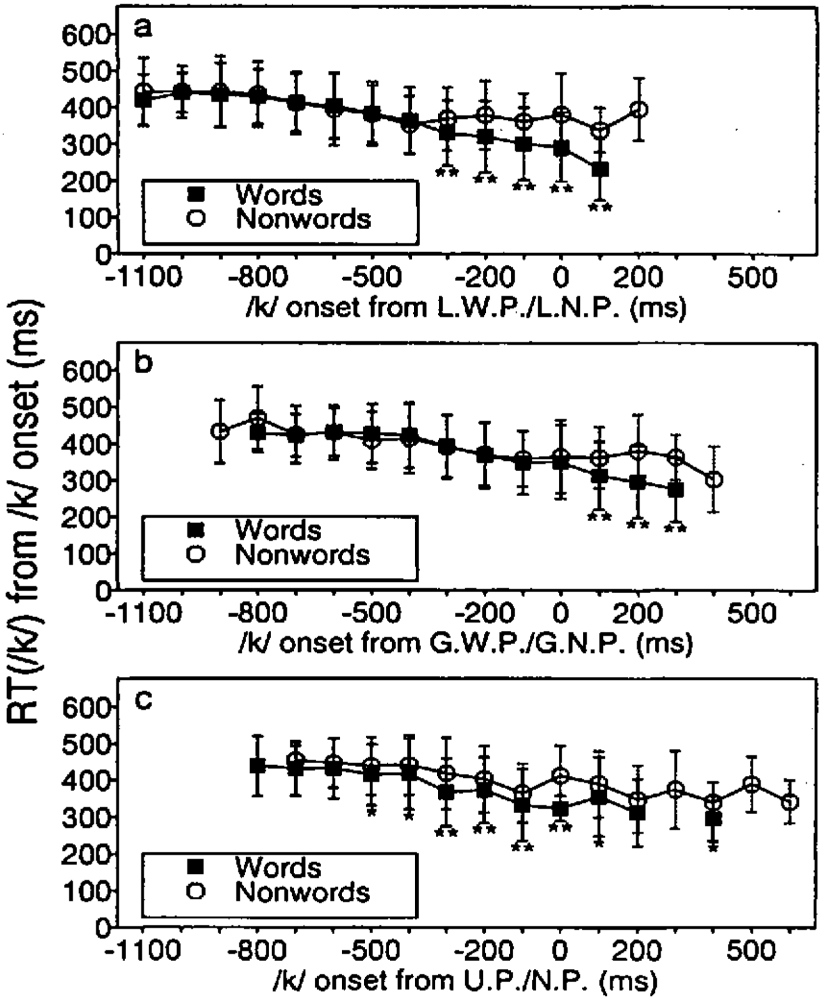


Figure 1: Mean reaction times for the detection of phoneme /k/, RT(/k/), as a function of the burst onset time of phoneme /k/, as measured from the reference point for word/nonword recognition estimated by (a) lexical decision task (L.W.P./L.N.P.), (b) gating task (G.W.P./G.N.P.), and (c) uniqueness point (U.P./N.P.). RT(/k/) is grouped into 100-ms slices along the horizontal axis. The bars and dashed bars represent the SE of words and nonwords, respectively. Single and double asterisks indicate a significant difference of 5% and 1% between words and nonwords, respectively.

The mean reaction times for phoneme /k/ in words and nonwords in Figure 1b differed significantly at 100 ms on the horizontal axis, $F(1,109) = 6.87$, $p < .01$, at 200 ms, $F(1,67) = 12.3$, $p < .01$, and at 300 ms, $F(1,29) = 8.67$, $p < .01$.

The mean reaction times for phoneme /k/ in words and nonwords in Figure 1c differed significantly at −500 ms on the horizontal axis, $F(1,304) = 5.15$, $p < .05$, at −400 ms, $F(1,247) = 4.15$, $p < .05$, at −300 ms, $F(1,345) = 24.3$, $p < .01$, at −200 ms, $F(1,279) = 7.25$, $p < .01$, at −100 ms, $F(1,445) = 14.7$, $p < .01$, at 0 ms, $F(1,22) = 11.5$, $p < .01$, at 100 ms, $F(1,226) = 5.13$, $p < .05$, and at 400 ms, $F(1,47) = 7.08$, $p < .05$.

## 4. DISCUSSION

The beginning point of the lexical effects is clear when the reference point for word/nonword recognition is estimated by lexical decision task (Fig. 1a) or gating task (Fig. 1b). The beginning point is unclear, however, when estimated by uniqueness point (Fig. 1c). This result conflicts with the assertion that the uniqueness point is the turning point for lexical effects ([4], [5]).

As Frauenfelder, Segui, and Dijkstra [5] discussed, the uniqueness point has some defects as the reference point for word/nonword recognition. First, the uniqueness point changes according to the size of the units in the dictionary. It is not known which unit size is perceptually reasonable and should therefore be used for calculating the uniqueness point. Second, the uniqueness point varies with the number of words in the dictionary, so it may differ from one subject to another. It is not known how many words the dictionary should have because of subject individuality. Because of these problems, the turning point for the lexical effects is not clear when the uniqueness point is used as the reference point for word recognition.

It may be that the unit size and number of words in the dictionary being used for the uniqueness point calculation in this experiment are inadequate. But even when the uniqueness point was calculated using a different unit size (phoneme/mora/syllable) and the number of words (60,000/5,000) in the dictionary, a clear separation of lexical effects was not observed. Calculating the appropriate uniqueness point is therefore not easy.

The uniqueness point problems might be solved if a dictionary which is identical to the subject's mental lexicon is used. But even under such ideal conditions, the uniqueness point may fail to show a clear separation of lexical effects because, as some studies (e.g., [7], and [11]) have shown, spoken word perception cannot always follow the left-to-right process on which the uniqueness point depends. The uniqueness point is therefore not suitable as the reference point for word/nonword recognition at which the lexical effects begin.

In contrast to the uniqueness point, the beginning point of the lexical effects is clear when either lexical decision task (Fig. 1a) or gating task (Fig. 1b) is used. The beginning points, however, are different: −400 ms for the former and 0 ms for the latter. Why don't they coincide?

This is because subjects have more time to respond in gating task than in lexical decision task. In gating task, they can precisely analyze the stimulus by using their memories and knowledge. They can therefore more easily perceive the word, even if only an initial-word fragment is presented. The G.W.P./G.N.P. probably represents the minimum stimulus length required for word/nonword recognition. And phoneme perception after the G.W.P./G.N.P. is facilitated by the lexical information.

In lexical decision task however, subjects have to respond as quickly as possible, so they do not have as much time to analyze the stimulus. They therefore need a longer fragment to perceive words or nonwords and reaction times for word/nonword recognition become longer than in gating task (the mean L.W.P. was 797 ms and the mean G.W.P. was 527 ms). Lexical decision task is a speeded task and reflects an ongoing perceptual process much more than in gating task. The present results indicate that lexical information facilitates phoneme perception even before a word is perceived and that the beginning point is 400 ms before L.W.P./L.N.P.

Of course, the reaction time for the lexical decision includes a neural delay between the decision being made in the brain and the corresponding muscle reaction in the subject's finger. If this neural delay were subtracted from the data, lexical effects before L.W.P./L.N.P. might disappear. According to physiological experiments on monkeys [16], this delay is about 100 ms, which is much shorter than the 400-ms delay between the beginning point of lexical effects and the L.W.P./L.N.P. Therefore, even if the neural delay is subtracted from the data, lexical effects before L.W.P./L.N.P. are still a factor in the ongoing perceptual process.

By the way, the isolation point (I.P.), at which a subject correctly guesses the word and does not subsequently change his or her guess ([10], [11]), is also obtained in gating task. The mean I.P. was 387 ms in this experiment. Interestingly, the 400-ms delay between the lexical effect beginning point and the L.W.P./L.N.P. is almost equal to the 410-ms delay between the I.P. (387 ms) and the L.W.P. (797 ms). This indicates that the lexical effect beginning point coincides with the I.P. The confidence rating is usually low at the I.P., so the subjects are still unsure of which word was presented. Some word candidates, including the correct one, probably emerge at I.P. and they may cooperationally facilitate the phoneme perception after that point in the ongoing perceptual process.

In summary, reaction times to the phoneme /k/ and the reference point for word/nonword recognition were measured to clarify their relationship. The stimuli were spoken Japanese words and nonwords of 1 to 5 mora long and contained one /k/ at various mora positions. The reference point for word/nonword recognition was estimated using a lexical decision task, a gating task, and a uniqueness point which was calculated using a Japanese dictionary of about 60,000 words. Results of the lexical effects on phoneme perception were not clear when the reference point for word/nonword recognition was estimated using the uniqueness point. The effects were clearly observed however in the lexical decision and gating tasks. The results show that the reference point for word/nonword recognition estimated using the lexical decision or gating tasks is better as a starting point for lexical effects than that estimated with the uniqueness point.

## 5. REFERENCES

[1] S AMANO, 'Effects of Lexicon and Coarticulation on Phoneme Perception', The Journal of the Acoustical Society of Japan (*in press*)

[2] W MARSLEN-WILSON, 'Function and Process in Spoken Word Recognition – A Tutorial Review', In H BOUMA & D G BOUWHUIS (Eds.), Attention and Performance: Vol. 10, London: LEA, pp. 125-150 (1984)

[3] U H FRAUENFELDER & J SEGUI, 'Phoneme Monitoring and Lexical Processing: Evidence for Associative Context Effects', Memory & Cognition, 17 134-140 (1989)

[4] U H FRAUENFELDER, J SEGUI, & T DIJKSTRA, 'Lexical Effects in Phoneme Monitoring: Facilitatory or Inhibitory?', Proceedings of the 11th International Congress of Phonetic Sciences, 105-107 (1987)

[5] U H FRAUENFELDER, J SEGUI, & T DIJKSTRA, 'Lexical Effects in Phonemic Processing: Facilitatory or Inhibitory?', Journal of Experimental Psychology: Human Perception and Performance, 16 77-91 (1990)

[6] B GORDON, 'Subjective Frequency and the Lexical Decision Latency Function: Implications for Mechanisms of Lexical Access', Journal of Memory and Language, 24 631-645 (1985)

[7] J C GOODMAN & J HUTTENLOCHER, 'Do We Know How People Identify Spoken Words?', Journal of Memory and Language, 27 684-698 (1988)

[8] L M SLOWIACZEK & D B PISONI, 'Effects of Phonological Similarity on Priming in Auditory Lexical Decision', Memory & Cognition 14 230-237 (1986)

[9] S COTTON & F GROSJEAN, 'The Gating Paradigm: A Comparison of Successive and Individual Presentation Formats', Perception & Psychophysics, 35 41-48 (1984)

[10] F GROSJEAN, 'Spoken Word Recognition Processes and the Gating Paradigm', Perception & Psychophysics, 28 267-283 (1980)

[11] F GROSJEAN, 'The Recognition of Words after Their Acoustic Offset: Evidence and Implications', Perception & Psychophysics, 38 299-310 (1985)

[12] A SALASOO & D B PISONI, 'Interaction of Knowledge Sources in Spoken Word Identification', Journal of Memory and Language, 24 210-231 (1985)

[13] L K TYLER, 'The Structure of the Initial Cohort: Evidence from Gating', Perception & Psychophysics, 36 417-427 (1984)

[14] L K TYLER & J WESSELS, 'Quantifying Contextual Contributions to Word-Recognition Processes', Perception & Psychophysics, 34 409-420 (1983)

[15] L K TYLER & J WESSELS, 'Is Gating an On-Line Task? Evidence from Naming Latency Data', Perception & Psychophysics, 38 217-222 (1985)

[16] E V EVARTS, 'Pyramidal Tract Activity Associated with a Conditioned Hand Movement in the Monkey', Journal of Neurophysiology, 29 1011-1027 (1966)