Frequency just above the second formant region, which is similar to the initial
(initial antecedent) of the third formant for Japanese, and the transition of their third formant. The
their initial steadier-rates and the transition of their third formant. The
differ as follows: English /l/ and /r/ are distinguished by differences in
japanese speakers. The English and Japanese videographs
lower approximation /l/ and higher approximation /r/. /l/ is unique to Japanese English pronunciation
like a gap or a consonant gap (Bock 1995). Thus, phonetically the gap-
are varieties have been difficult to describe articulatorily, but it appears to be
varies. While Japanese has only one liquid pronunciation, /r/, and
level. Therefore, the English approximation /l/ and /r/ are
pronounced in Japanese. Pronunciation in Japanese is that it allows no opaque production
pronunciation in Japanese. Pronunciation in Japanese is that it allows no opaque production
language by Japanese learners is that it allows no opaque production in
language by Japanese learners. This is of interest in studying the acquisition of English
pursue, pronunciation and production of English /l/ and /r/ for Japanese learners. One set of studies has
clear differences in the pronunciation of Japanese speakers. These differences provide no support for the
language by Japanese speakers. These differences provide no support for the

The study examines the role of duration in the perception of a new

Elianshel Henry

University of Minnesota

AND JAPANESE SPEAKERS

CANTONESE

OF ENGLISH AND A COMPARISON OF

PERCEPTION

DURATION AND CONTEXT EFFECTS ON THE
In an attempt to understand why English word-initial positions tend to be relatively easy, they examined English English in world-final position. The results suggested that English words in world-final position were easier to recognize than those in world-initial position.

The performance of Japanese speakers on world-final English was superior to that of world-initial English. This was consistent with previous research that showed Japanese speakers have an advantage in recognizing English words in world-final position. The results also suggested that the initial consonant position is a significant factor in determining the difficulty of recognizing English words in world-final position.

In conclusion, the study demonstrated that the initial consonant position has an important effect on how well Japanese listeners recognize English words in world-final position. This suggests that the phonological structure of English words plays a role in their recognition.
Speakers of Japanese have the opportunity to perceive the acoustic signal that results from a phonological contrast, which is not present in either native or non-native speech. The results of Dissonant-Height et al. (1982) suggest that this is true when native speakers of English, who are similar to what would be expected of native speakers of Japanese, are given a language and asked to discrimination of these sounds in final position was quite tight with a phonological contrast, which is not present in either native or non-native speech. A native speaker of English would presumably have lost or shied away from identifying the English Nulls in its full and complete form. In this study, we were not interested in whether native speakers can be accounted for on the basis of native language experience; while the Japanese speaker was able to detect the acoustic properties on a phonetic basis, the native speaker was not able to do so. The Duration Hypothesis states that Japanese sub-lexical units of perception are in the Duration Hypothesis.

<table>
<thead>
<tr>
<th>8</th>
<th>10</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>1</td>
<td>F</td>
<td>M</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>14</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>11</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Combined</td>
<td>/1/</td>
<td>/1/</td>
</tr>
</tbody>
</table>

**Word Position**

<table>
<thead>
<tr>
<th>11</th>
<th>11</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>14</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>17</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>11</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Combined</td>
<td>/1/</td>
<td>/1/</td>
</tr>
</tbody>
</table>

**Consensus Cluster**

<table>
<thead>
<tr>
<th>I</th>
<th>11</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>14</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>17</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>11</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Combined</td>
<td>/1/</td>
<td>/1/</td>
</tr>
</tbody>
</table>

**Final Cluster**

---

*Table 1: Dissonant-Height et al. (1982)*

---

Henry and Sheldon
English phonetic approximation. The distribution of /l/ in English is a continuum, ranging from a voiceless, unaspirated sound to a voiced, aspirated sound. In English, the /l/ sound is produced by rounding the lips and the tongue, creating a frictional sound.

In contrast, the /l/ sound in another language, such as Spanish or Italian, is produced differently. Spanish speakers, for example, produce the /l/ sound with a velarized articulation, whereas Italian speakers produce it with a labialized articulation.

This difference in phonetic approximation can have implications for how speakers of one language perceive and learn the /l/ sound in another language. For example, Spanish speakers may have difficulty learning to produce the /l/ sound in English, as they are accustomed to producing it differently. Similarly, Italian speakers may have difficulty learning to produce the /l/ sound in Spanish, as they are accustomed to producing it differently.

Overall, the phonetic approximation of /l/ in different languages serves as a useful tool for understanding the differences in speech production and perception across languages.
Figure 1: Variation in formant frequencies of two English speakers' productions of /i/ in various vowel contexts.
Speakers of English

Cantonese and they had minimal contact. If any, with Cantonese
years old who was a senior in high school. None had had any exposure to
undergraduate students at the University of Minnesota, except the 17
they ranged in age from 17 to 27 years. All were either freshman or
lived in dormitory for four years after high school, at least. The four American
women and one man), they were natives of Minnesota, the fourth and
OF THE FOUR AMERICAN SUBJECTS WHO PARTICIPATED IN THIS STUDY (ONE

Problem)

percent of the time. They were at an advanced level of English
of the study they estimated that they were conversational in English at least
instruction in Hong Kong from a native speaker of English. At the time
eight years of age. All except one had received some higher education
years of age, and had lived in the United States from 1 1/2 to 4 1/2 years,
and three men, who were undergraduate students enrolled in academic
and Cantonese speakers (Cantonese, two women
occupation and the results of
received in the results of
Sheldon 1983). The data pertinent to this study are the results of
and /æ/ with which is discussed in greater length elsewhere (Fry 1976) and
study of Cantonese speakers pronunciation and production of English
and /æ/ and /æ/ were examined here because whereas in another
as in the pronunciation and production of English
and /æ/ the production data examining the ability of Cantonese speakers to identify English
and /æ/ discussion of the ability of Cantonese speakers, to identify English
and /æ/ differences between the English and Cantonese bilinguals will turn to a
the northern and second forms raised in English

CANTONESE SPEAKERS’ PERCEPTION OF ENGLISH LANGUAGES

selected /æ/ of English.

examined exhibited the formal pattern associated with the nasal,
in the English. None of the Cantonese varieties of /æ/ and /æ/ in the Cantonese
in a relatively greater distance between /æ/ and /æ/ in the English
formant of the English /æ/ in comparable vowel contexts. This results in
second formant of the English /æ/ in the English words.
In fact, the second formant of the English /æ/ in the English words
has a higher third formant and the first and second formants are relatively

Language Learning...
was: world-initial, world-medial, world-final; consonant cluster.

English words are a function of position, ranked from easiest to hardest, word-initial, world-medial, world-final. Consonant cluster.

Several factors have been identified as affecting the perception of English words, including position in the word, initial and medial consonant clusters, and the presence of a consonant cluster. These factors have been found to influence the perception of English words, with position in the word being the most significant factor. Position in the word, initial and medial consonant clusters, and the presence of a consonant cluster all contribute to the perception of English words.

The results were confirmed by the statistical analysis, which showed a significant difference in the perception of English words, with position in the word being the most significant factor. Initial and medial consonant clusters also played a significant role in the perception of English words. The presence of a consonant cluster further enhanced the perception of English words.

METHOD

| RESULTS |

| Headline and Sheldon | 31 |

Sequence included 96 randomized stimuli: Two repetitions of the 48 test words were also included in the stimulus material. The complete test position followed the same procedure as Sheldon and Shane (1989). The Chinese subjects were given two forced-choice tests. The average for these tests was 76% correct.
Measurement of the Duration of the Liquids

One of our data sets revealed that the word-final liquids in Japanese are shorter than the word-final liquids in English. This was predicted by the hypothesis that the duration of the word-final liquid would be longer than the word-final liquid in English. However, this prediction was not supported by the data collected.

In addition, the specific comparisons between the means (Wilcoxon matched-pairs signed-ranks test) showed significant differences for the liquids in fricative-initial position. The liquids in English were longer and more consistent than the liquids in Japanese.
We then computed the durations of the high consonants in the words

The durational effect was significant, with a mean duration of 127 ms, the duration of the high consonants in word-final position, ranging from 70 to 111 ms. The duration of the high consonants in word-initial position was 188 ms, with a mean duration of 127 ms. The duration of the high consonants in word-medial position was 188 ms, with a mean duration of 127 ms. The high consonants in word-initial position were the least well-identified.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Initial</th>
<th>Final</th>
<th>Medial</th>
<th>Overall</th>
<th>Combined Mean</th>
<th>Combined Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>149</td>
<td>149</td>
<td>149</td>
<td>149</td>
<td>149</td>
<td>149</td>
</tr>
<tr>
<td>2</td>
<td>129</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
<td>132</td>
</tr>
<tr>
<td>3</td>
<td>111</td>
<td>112</td>
<td>112</td>
<td>112</td>
<td>112</td>
<td>112</td>
</tr>
<tr>
<td>4</td>
<td>109</td>
<td>110</td>
<td>110</td>
<td>110</td>
<td>110</td>
<td>110</td>
</tr>
<tr>
<td>5</td>
<td>98</td>
<td>101</td>
<td>101</td>
<td>101</td>
<td>101</td>
<td>101</td>
</tr>
<tr>
<td>6</td>
<td>88</td>
<td>94</td>
<td>94</td>
<td>94</td>
<td>94</td>
<td>94</td>
</tr>
<tr>
<td>7</td>
<td>86</td>
<td>90</td>
<td>90</td>
<td>90</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td>8</td>
<td>68</td>
<td>68</td>
<td>68</td>
<td>68</td>
<td>68</td>
<td>68</td>
</tr>
</tbody>
</table>

Table 2:

Henry and Sheldon.
Identification errors occurred above the mean and fourteen errors duration of the liquid in consonant cluster position is 12 ms. Fifteen errors are also spread out across the range of duration. The mean, not cluster in the region of short duration. In consonant cluster position, and can errors occur with a duration below the mean. The errors do identification errors on liquids with a duration that was above the mean. Identification errors on liquids with a duration that was below the mean.

![Graph](image)

**Figure 2.** Duration (in milliseconds) of English liquids for each token in the native speakers of Cannae.

Results of a forced-choice identification test performed by 5 native speakers. The perception data (represented by X) are position. The perception data (represented by O) are the horizontal axis. The tokens shown below the mean duration of liquids in each word position. Sectors of each word position are colored according to the percentage of errors in each word position. (Error bars are calculated from standard deviations of 27 tokens.)

<table>
<thead>
<tr>
<th>Word Position</th>
<th>Initial</th>
<th>Medial</th>
<th>Final</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>xx</td>
<td></td>
<td></td>
<td></td>
<td>00</td>
</tr>
<tr>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>00</td>
</tr>
<tr>
<td>xx</td>
<td></td>
<td></td>
<td></td>
<td>00</td>
</tr>
<tr>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>00</td>
</tr>
<tr>
<td>xxx</td>
<td></td>
<td></td>
<td></td>
<td>00</td>
</tr>
<tr>
<td>o</td>
<td></td>
<td></td>
<td></td>
<td>00</td>
</tr>
<tr>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>00</td>
</tr>
</tbody>
</table>

Henry and Sheldon
and since English /l/ in final position is quite different from
Caninese speakers are using their own /l/ as the practical "template" for
decide that their /l/ is not equivalent to English /l/ and /l/ could cause the Caninese speakers to
not hear similarity in acoustic structure between Caninese and
and the relative pronunciation of the English /l/ in non-final word positions. The
earmark the template is identified as /l/, while those
those liquids that are marked "template" are identified as /l/. Thus, their
Caninese /l/ as a template for judging all English /l/ variants.
their Caninese /l/ is a template for judging all English /l/-variants.
Caninese /l/ is a template for judging all English /l/-variants.

The Duration Hypothesis predicts that, all other things being equal,

English liquids in word-final position should be more easily perceived by
English listeners than liquids in word-initial position due to the relatively long duration of the acoustic

All three of the previously-mentioned studies of Japanese listeners of

DISCUSSION

The Duration Hypothesis predicts that, all other things being equal,

English liquids in word-final position should be more easily perceived by
English listeners than liquids in word-initial position due to the relatively long duration of the acoustic

All three of the previously-mentioned studies of Japanese listeners of

English liquids in word-final position should be more easily perceived by
English listeners than liquids in word-initial position due to the relatively long duration of the acoustic

All three of the previously-mentioned studies of Japanese listeners of

English liquids in word-final position should be more easily perceived by
English listeners than liquids in word-initial position due to the relatively long duration of the acoustic
The word-initial cluster was either voiced or voiceless. This makes com-

discourse-prints et al. (1982) Japanese study the initial stop consonant in

plantar and a labial, whereas in the English study of Sh dilation and enlarge-
cセンターに示されると clustering of a voiceless stop con-

be more complex to measure in

Returning to our results in consonant cluster position, we found that

were acquired by the greater degree of phonetic similarity between their

was lower for all clusters than that of the Japaneses. Consequently, percep-

tion in word-initial and word-medial position consonant clusters had

English while Japanese does not. Actually measured in the consonant

earlier language can influence the degree of difficulty in learning and

This comparison of the Japanese and Canaunose results reveals how

in Table 1). In fact, with the form of /l/, /l/ would be accurate and most mistakes would be made on /l/.

This problem is accentuated by the fact that the Canarinese speakers had

uttered /l/, /l/ and /l/ in English, one would expect that Canarinese speakers would be

Given the above explanation of how the Canarinese speaker identifies

light. Given the above explanation of how the Canarinese speaker identifies

Japanese speakers should be able to make advantages of the longer

Japanese speakers, there is no reason to assume that Japanese speakers could

and hence no peculiar representation which is similar to the English

On the other hand, Japanese speakers have no native language phone

the native language.

The effect of duration on perception is thus less important than the influence of

words, the greater duration of the English light. By itself, it is not suffi-

amount of time should not affect their identification judgments. In other

Canarinese /l/ is, however, the more opportunity to perceive the English /l/ for a longer

Henry and Sheldon

17
Language Phoneme Perception.

Conversational factors that are affecting the process of second-language phoneme perception are attesting the process of second-language phoneme perception. Speakers with different native language phonology perceive prosodically similar English sounds while other languages. The development of English phonology, as described by Japanese speakers, is not predicted by Japanese phonology. Since the consonant clusters in Japanese are not produced by Japanese phonology, this conclusion is not surprising. The phonology of Japanese has an influence on the phonology of the consonant clusters in Japanese. However, this influence does not apply to the phonology of Japanese. Therefore, the consonant clusters are not affected by the phonology of Japanese.

Duration did not facilitate perception. For these reasons, we attempted to find the limits in consonant clusters where the duration of these clusters was longer than some of the single liquids in other positions. The single liquids are significantly longer than the consonant clusters. In addition, the consonant clusters were shorter than the single liquids in the previous positions. Information about consonant clusters because of the effect of coordination is produced a more complex speech signal to encode than single consonant clusters. Information in the voiceless part of the signal, consonant clusters, and information in the voiced part of the signal, voiceless clusters, are encoded in addition to the loss of short, high percentage of the signal. Secondary to the coordination is a second position during the transition of the high pitch to the following vowel. This makes the transition of the high pitch to the following vowel—particularly the higher quality portion of the signal—in voiceless clusters after the voiced portion of the signal. The transition to the following vowel is more difficult, for the following reasons: First, the lower amplitude of the voiceless clusters contribute to the consonant clusters, which is encoded in the following syllable. However, we suggest that other variables besides duration, such as segmentation of the consonant clusters in these cross-language studies.

Language Learning, Vol. 36, No. 4
We cross-checked the reliability of our measurements by designing some of our stimuli sample to have multiple possible correct answers, where the correct answer was more probable to be a correct answer. For these, we could not make a clear-cut decision on the ideal position measurement correlation to the sound of the steady-state portion of the stimulus. In one example, the measurement of the ideal position and the sound of the steady-state portion of the stimulus were coincident at the beginning of the sentence. In another example, the measurement of the ideal position and the sound of the steady-state portion of the stimulus were coincident at the end of the sentence. In these examples, the precision of the measurement of the ideal position and the sound of the steady-state portion of the stimulus correlated to the beginning or end of the sentence.

American

There was a slight loss of the stimulus in each of the four word positions, spoken by four native speakers of English. The listeners were instructed to indicate the position at which they felt the stimuli to be the most intelligible, and to indicate the position at which they felt the stimuli to be the least intelligible, and to indicate the position at which they felt the stimuli to be the least intelligible.

The authors' names are listed alphabetically. Further versions of this paper were published in.

NOTES

The results of our study indicate that native English speakers have a more accurate perception of English phonemes than do native speakers of other languages. This is because English speakers are more familiar with the sounds of English, which are more consistent across different speakers and situations. The results also indicate that cross-linguistic listeners are more accurate in their perception of English phonemes than are native English speakers.

In conclusion, we have shown that native English speakers have a more accurate perception of English phonemes than do native speakers of other languages. This is because English speakers are more familiar with the sounds of English, which are more consistent across different speakers and situations. The results also indicate that cross-linguistic listeners are more accurate in their perception of English phonemes than are native English speakers.

Henry and Sheldon

519
on a PDP 11/34 computer. The resulting measurements showed no appreciable difference in accuracy over the measurements from the spectrograms.

4For example, on some spectrograms there was no formant structure during the voiceless portion of the cluster, where one would expect formant structure associated with the liquid. There was, instead, aperiodic acoustic energy.

It is possible that English /l/ in consonant cluster position can be somewhat velarized, which would predict a response bias toward /l/ according to our hypothesis of how Cantonese speakers perceive English liquids. We checked the spectrograms for consonant cluster /l/ and found no evidence of velarization. This would show up as a greater degree of proximity between the first and second formant than would normally be present for the non-velarized variant of /l/. Since the beginning of the liquid in word-initial consonant cluster position was voiceless and the amount of acoustic energy was so low in the period immediately following the voiceless stop burst, no formant structure was apparent. What was evident was either aperiodic energy (aspiration) or no noticeable acoustic energy at all. The one stimulus that exhibited clear formant structure in the consonant cluster was play. The first and second formant positions for /l/ in play were comparable to those of the initial /l/ in the stimulus lay. F1 and F2 of the /l/ in play were 286 Hz. and 733 Hz., respectively; in lay F1 and F2 of the /l/ were 250 Hz. and 715 Hz., respectively. Thus, it can be assumed that the English consonant cluster /l/ in our data is like other English non-final liquids. This would facilitate the Cantonese subjects' perception of English /l/ in word-initial consonant clusters.

If such a correlation were to be found, then the next step would be to experimentally manipulate duration in order to see if longer duration does in fact increase the accuracy of identifying /l/ and /l/. If there is a causal relationship, then one practical application could be to train Japanese learners on speech samples in which the duration of the liquid has been lengthened. This would be a reasonable approach to the problem of simplifying this difficult perceptual learning task. In an attempt to make a preliminary comparison between the duration measurements reported by Dissosway-Huff et al. (1982) with those for the English word stimuli that were used in the Japanese study of Sheldon and Strange (1982), we measured the duration of the liquid consonants for one of the four American speakers in Sheldon and Strange (1982). Measurements were taken of all of this speaker's tokens of /l/ and /l/ in each of the four word positions. This gave a total of 32 tokens, eight in each of the four word positions, half of which were measurements of /l/ and half of /l/. The resulting combined mean durations for /l/ and /l/ were similar, relatively speaking, to the ones found by Dissosway-Huff et al. (1983). Our combined mean durations for /l/ and /l/ were as follows: word-initial position, 86 ms., consonant cluster, 106 ms., word-medial, 100 ms., word-final, 171 ms.

REFERENCES

APPENDIX A

Hendy and Sheldon