Speech Rates and a Word Recognition Ratio for Listening Comprehension of Movies

NITTA Haruhiko (Senshu University)
OKAZAKI Hironobu (Akita Prefectural University)
Walter KLINGER (University of Shiga Prefecture)

1. Introduction

One particular objective of many people who study English as a foreign language is to understand English language movies without having to resort to reading the subtitles, in the same way one can enjoy movies in one’s native language. However, to be able to follow the storyline not even totally but just to a satisfactory degree is a goal difficult to achieve. Why is it so difficult to hear well? Understanding what is being said in a movie involves a number of intellectual factors, such as knowledge of vocabulary, idioms, and grammatical structures, and knowledge of social and cultural settings, which is particularly needed to understand humor. We believe the most important and basic ability for listening is to be able to “decode,” to hear and identify, the sounds of the words being spoken. Unless you can decode what is heard, your intellectual knowledge of written words does not help much. Moreover, the decoding ability of non-native speakers of English (NNSE) is likely to be impaired or lost when native speakers speak fast. In this article, we look at a number of research studies about speech speed and its effect on listening ability, determine a Word Recognition ratio for the percentage of words that NNSE can decode in movies, and discuss why it is so hard for NNSE to decode what is spoken in movies.

2. Reports of Research Studies on Speech Speed and Listening Ability


A number of research studies have focused on listening comprehension as it is affected by speech speed. Wingfield, Lombardi & Sokol (1984) used magazine articles and essays to study native speakers’ listening comprehension, and reported that the number of words accurately heard showed a decrease as the speaking speed increased. Griffiths (1992) used listening material at slow, average, and fast speeds with lower-intermediate Japanese elementary school teachers, and reported a similar result. Tomita (1998) used ordinary, fairly slow, and very slow listening material with Japanese first-year university students, and reported that increased speech speed affects comprehension. These studies all confirm the general impression one gains from one’s own experience that speech speed affects listening ability.

2.2 Nitta, Okazaki & Klinger (2010b)
Nitta, Okazaki & Klinger (2010b) investigated the relationship between listening ability and different rates of speech speeds, using 60 sentences from American TV shows (Friends, Sex and the City, and Beverly Hills 90210). The sentences were divided into groups of 12 sentences spoken at 5 different rates of speed, measured in syllables per second (sps). All of the subjects were allowed to listen to the source material any number of times until they were satisfied with what they could hear, and then record by typing what they heard. A few words unknown to the NNSE in written form were eliminated from the results.

The subjects were 31 non-native speakers of English (NNSE) and an equal number of native speakers of English (NSE). The NNSE all had over 860 points in their TOEIC® scores, with an average score of 923.3, and had an average of 3.5 years of overseas residence in English-environment countries. Twenty of them had an average of 4.6 years of undergraduate and post-graduate study in universities in English-environment countries, and many of the other eleven had overseas experience, though not formal overseas school experience. The subjects had apparently thorough and sufficient experiences in English language study and practice, and a high level of evaluation in English ability. The results of the experiment are shown in Figure 1.

Figure 1. Missed Word Rates (Nitta, Okazaki & Klinger 2010b).

At 4 sps, the NNSE subjects missed or mistook 4.2% of the words, of which 2.7% were function words and 1.5% were content words. At 5 sps, the Missed Word Rate (MWR), the number of error words (missed or mistaken words), jumped to 12.6%; 10.5% function words and 2.1% content words. At 6 sps, the MWR was 21.2%; 16.4% function words and 4.8% content words. At 7 sps, the MWR was 32.7%; 24%...
function words and 8.7% content words. At 8 sps, the MWR was 40.6%; 30.1% function words and 10.5% content words. Interestingly, at 7 sps and 8 sps, the NSE subjects also began to miss words.

2.3 Nitta, Okazaki & Klinger (2010a)

In an earlier study, Nitta, Okazaki & Klinger (2010a) analyzed 11,877 sentences from 11 movies and 3 episodes of TV series to determine articulation rates (AR), i.e., the speaking rate of speed in a unit of speech minus any long pauses, as measured in syllables per second (sps). The film media that were analyzed and their median AR in sps are as follows: three “human drama” films (A Beautiful Mind 5.4 sps, The Hours 4.9 sps, Out of Africa 4.9 sps), three classics (Roman Holiday 4.7 sps, The Seven Year Itch 5.1 sps, The Apartment 5.8 sps), two comedies (Back to the Future 5.3 sps, Legally Blonde 5.3 sps), two animation films (Ratatouille 4.3 sps, Monsters Inc 4.6 sps), one action film (The Bourne Identity 5.2 sps), and three TV dramas (Sex and the City 5.0 sps, Full House 4.9 sps, Columbo 5.3 sps). The AR of all 14 media combined are shown in Figure 2.

![Figure 2. Speech Speeds in Movies (Nitta, Okazaki & Klinger 2010a).](image)

The distribution of the AR in each movie has a mountain-like shape, indicating that people talk in various speeds from around 2 sps to over 8 sps, and the most frequent speed is around 5sp. Movies are often recommended for English language study because they are considered to be “authentic” material (e.g., Takahashi 1995; Kadoyama 2005). The popular notion that movies reflect authentic speech and authentic speech speeds suggests that Figure 2 also reflects the everyday speech speed patterns in conversation of native speakers. Robb et al. (2004) list a number of research works showing that adult speakers of American English have an overall AR of about 5.3 sps.
We have taken the data of Figure 2 and shown it in a percentile in Figure 3.

![Figure 3. Movie Dialogue Speed as Percentiles (Based on Nitta, Okazaki & Klinger 2010a).](image)

The chart shows that 18.7% of a typical movie scenario is spoken at up to 4 sps, 48.3% is spoken at up to 5 sps, 77.2% at up to 6 sps, 93.5% at up to 7 sps, and 99.2% at up to 8 sps. The chart clearly tells us that we need to cope with faster speeds in order to obtain more complete decoding in listening.

3. Word Recognition Ratio for Movies

3.1 Deriving a Word Recognition Ratio from Figure 1

Figure 1 showed that the faster the speech rate, the worse the Missed Word Rate. Figures 2 and 3 showed that any movie has fast speaking parts to some extent. Then what happens when NNSE see movies? How much of the speech can they decode? We would like to take the results of Nitta, Okazaki & Klinger’s two studies (2010a and 2010b), and calculate the Word Recognition Ratio (WRR), the complementary idea of the Missed Word Rate, to see how much of a full movie they could theoretically understand.

First, we look at Figure 2 from another distribution ratio. 8.1% of all movie speech is up to 3.4 sps. 21.5% is between 3.6 sps and 4.4 sps. 31.6% is between 4.6 sps and 5.4 sps. 23.9% is between 5.6 sps and 6.4 sps. 11.5% is between 6.6 sps and 7.4 sps. And only 2.9% is between 7.6 sps and 8.4 sps.

We apply Figure 1 data to these figures to calculate the total WRR.

\[
\begin{align*}
4 \text{ sps} & \quad 4.2\% \\
\times & \quad 21.5\% \\
= & \quad 0.9\%
\end{align*}
\]

4 sps  \times 21.5\% = 0.9%
We obtain a Word Recognition Ratio (WRR) of 85.1%. We ignore the 8.1% of the movie composed of speech of up to 3.4 sps since we have no MWR data under 4 sps. If we had any MWR data under 4 sps, it would make the WRR worse than 85%. 85% might be a rough figure, but it gives us information that we can apply later.

For the NSE, we calculate that they would have a WRR of 99.4%. Field (2009, p. 289) reported on a word identification test that demonstrated that, for native speakers, “Their accuracy with most words was 100%, and it did not drop below 96%.” Our calculation is in line with Field’s report.

3.2 Word Recognition of university students

We determined that 85.1% is the WRR for advanced NNSE. We would also like to see the WRR of ordinary university students. Okazaki & Nitta (2005) used sentences spoken by a native English speaker at a normal speed (in the same manner as used to speak to other native speakers), and reported the results of a listening test of 270 Japanese university students, as shown in Figure 4.

![Figure 4. Recognition Ratio of Words Heard (Okazaki & Nitta 2005).](image)

Although all of the subjects recognized all of the vocabulary in written form, in the listening test most
subjects recognized only 40-65% of the words. The subjects who recognized 90-95% were foreign exchange students, not Japanese college students. The Japanese subjects were first year college students whose English level was typical; i.e., their TOEIC® score would probably be well under 500 points (incoming students at one particular public university, where all students take the test, average about 346-370). These subjects typically had weaknesses in recognizing personal and common pronouns, auxiliary verbs, prepositions, words following words ending in plosives, linked words, and sounds spoken without stress.

3.3 Word Recognition vs Comprehension

We calculated the Word Recognition Ratio of the NNSE in Figure 1 to be 85%, but we cannot confidently say with the present data how much of a movie a NNSE with a WRR of 85% can actually understand. Nitta, Okazaki & Klinger (2010b)’s experiment evaluated listening ability to “decode” the sounds, rather than listening “comprehension.” The experiment recorded only the number of error words; it did not evaluate the subjects’ comprehension of the heard sentences and it did not determine how much comprehension decreases as speech speeds increase.

Listening comprehension is affected not just by hearing the actual words in a given sentence, but by understanding the context of the sentence, as it relates to what was previously said or its social setting, in a “top-down” process (Goodman, 1970). This kind of guessing or assuming is useful in comprehending meaning when only a few words are missed or mistaken; but, as more and more words are missed or mistaken, the meaning of the communication is also surely misunderstood despite attempts at top-down processing.

As for “bottom-up” processing (Gough 1972), listening to the smallest phonemes or syllables in a word and building them up to form a complete word and then words into phrases, it is often supposed that content words are more important for understanding than function words, i.e., that, in the phrase “a dog,” “dog” is more important than “a.” Nonetheless, the large number of function words missed by the subjects in Figure 1 poses the question of how their comprehension was affected by not hearing or catching function words. Field (2009, p.146) says,

Function words present an interesting case. On the one hand, they are very frequent in everyday speech. Of the 100 most frequent spoken items in the British National Corpus, a 100-million-word corpus of speech and writing, about 80 are function words. Surely, by dint of constant exposure, the listener must learn to identify them with a considerable degree of accuracy.

Field states the importance of function words, but for further arguments, more research is needed to identify to what degree comprehension is affected by missing function words as compared to missing content words.

To speculate what the Word Recognition Ratio of 85% means, the following studies might help. Laufer
(1989, p.319) says, “...reading academic prose is likely to be greatly affected by the lexical knowledge of the text. A chance to become a ‘reader’ is significantly higher if the lexical coverage of the test is 95% and above.” Nation (2001, p114) says “Learners would need at least 95% coverage of running words in the input in order to gain reasonable comprehension and to have reasonable success at guessing from context. A higher coverage of around 98% would be better.” On a later page, Nation again says that, “To gain adequate comprehension of the text, most readers would need to know 98% of the running words in the text. At the 95% density level, most learners did not gain adequate comprehension.”

Following Laufer and Nation’s arguments, a WRR of 85% is not at all adequate for the comprehension of movies. Moreover, the 85% figure is derived from an experiment where subjects were allowed to repeatedly listen to sentences (the NNSE listened an average of 5.2 times at 4 sps, 7.7 times at 5 sps, and 10.6 times at 8 sps), and where the subjects were familiar with all the words on paper. This figure of recognizing 85% of the words of a movie cannot be applied to a person watching a movie that must inevitably have unknown words, and watching that movie at a single viewing.

When we refer to Figure 3, we can see that in order to meet Laufer’s 95% coverage, we need to be able to decode speech at speeds a little bit faster than 7 sps, and if we try to meet Nation’s 98% coverage, we need to understand speech at almost 8 sps.

3.4 Subjects’ Self-Evaluation of Movie Comprehension

We executed a questionnaire survey among 29 of the NNSE subjects who participated in Nitta, Okazaki & Klinger (2010b)’s experiment, asking how much of a typical movie they thought they understood. Figure 5 shows the results of the questionnaire.

![Figure 5. NNSE Self-Evaluation of Movie Comprehension](image-url)
Six subjects replied that they understood 85% or more, which seems to mean being able to understand much of the movie in the original language. Half of the total subjects said they understood 50-60%, which seems to indicate at least some difficulty following the storyline. As these subjects had high TOEIC® evaluations, language learners with lower scores certainly would have much lower results.

4. Discussion

4.1 What seems to be the problem?

The advanced NNSE in Nitta, Okazaki & Klinger (2010b), who had a WRR of 85.1%, had a seemingly adequate educational background in English language studies, and their motivation to learn was high, as they often listened to daily English news and watched many movies. Yet, while they knew all the words on paper in the experiment, these subjects still missed or misheard 13% of the words at 5 sps and 21% at 6 sps. One of the possible explanations to this phenomenon is that the sounds were changed to the extent that the subjects could not decode them.

4.2 Variant Pronunciations affecting Listening Ability

Laver (1994, p.67) points out that pronunciation changes with speech styles, noting that, for example, the word “actually” can be pronounced in 7 different ways. Variations in pronunciation of a word can be due to regional dialects, formal or casual pronunciations, or due to expressing emotions like sadness or anger. “Speech style relies on three different types of the manipulations of the speech material of the utterance: re-organization of the phonemic structure of individual words, modifications of speech rate, and associated prosodic changes of pitch and loudness behavior,” says Laver (1994, p.67), and continues, “Polysyllabic words in English have on average about two or three such reorganized pronunciations for use in informal speech, though some have very many more alternatives.”

As speaking rates increase, pronunciation also changes. At very fast speeds, “errors are frequent and speech is distorted” (Laver 1994, p.543). Field (2009, p.143) notes that the pronunciation of words often differ when they are in connected speech and when they are said in isolation.

Variables in pronunciation certainly can make NNSE confused, and it might be the cause of some of the problems encountered by the NNSE subjects in Nitta, Okazaki & Klinger (2010b). Getting used to variables in pronunciation requires a lot of exposure to hearing language, and it is unrealistic to expect language learners, even those who have several or many years of residence in the target language environment, to be able to have as much exposure as native speakers. Indeed, some of the NSE subjects in that experiment also had some problems listening to high speaking rates.

Speech speed and variables in pronunciation, and how it affects listening skills of NNSE, is an area in
second language acquisition studies that needs more research.

4.3 What is a possible solution?

Usually, when using movies as learning resources, subtitles or scripts are used, and they are very useful when there are words or phrases that are hard to catch. According to our experiences as English teachers in classes for Japanese students, students have a tendency to be satisfied when they see the script and check the right word(s). They rarely take one more step and try to find out the reason why they could not decode what was spoken. When learning a spoken language, noticing and becoming familiar with its sounds is indispensable, but relying too much on subtitles or scripts might impair the interest or sensitiveness for the real sounds, and cost learners what’s really important.

Of course, it is the job of the teacher to teach students how to learn to distinguish and recognize sounds. But if the subjects of Nitta, Okazaki & Klinger (2010b) would ask us what they should do to improve their MWR at fast speeds, we could not be of any help, as we do not know the right method, either. Furthermore, while shadowing and repetition training are often used in language studies, we are not sure if they are effective for speech at 6 sps or faster.

Nonetheless, we tend to think that the ultimate way for improving listening skills may be to get the sounds that come out of your mouth close to the sounds that come into your ears. In other words, train yourself so that you can produce native-like sounds. Walker (2002, p166) recommends voice training and says “You will acquire genuine listening ability by using the sound detection and feedback system within your own body, so that supportive materials such as scripts or subtitles are no longer necessary.”

Listening ability seems to improve through interaction with speaking (e.g., Klinger 2001; Richards 2001; Celce-Murcia 1996). Most textbooks for language study these days contain practice conversations so that students will improve listening and speaking interactively (e.g., Hedberg & Mauser 2008; Brown 2007; 2006; Rolton 2005).

5. Conclusion

We looked at a study that showed that the dialogue between native speakers in movies is spoken at an average of about 5 syllables per second, but also showed that there are always some much faster sections. Another study verified the common observation that even advanced non-native speakers of English (NNSE) have difficulty in catching words at fast speech rates. From the data of two studies, we determined a Word Recognition Ratio of 85%, the percentage of words that advanced NNSE would be able to decode in a movie, given opportunities for repeated listening. Two researchers have said that at least 95% of words need to be known for reasonable comprehension. 93.5% of the content of movies is spoken at rates of speech up to 7 sps, but advanced NNSE missed 21.2% of the words spoken at 6 sps, 32.7% at 7
sps, and 40.6% at 8 sps. When watching movies, NNSE have to deal with speech speeds that are far beyond their ability, and we are not able to give assured answers to their problem at the present time.

We must point out that people can typically understand speech spoken at much higher speeds than is usually used for speaking (Laver 1994, p.543), so it should not be considered an insurmountable problem. We hope to continue with research on how to solve the problem and find new solutions to improve listening ability at high speeds.

Acknowledgments

This research was partially supported by a Grant-in-Aid for Scientific Research (B)(20320084) by the Japan Society for the Promotion of Science (JSPS).

References


Abstract

We examine previous research studies on listening ability and comprehension as it relates to speech speed; in particular, one study that determined that the average speech speed of sentences in English movies was 5.1 syllables per second and that 93.5% of the speech content was spoken at rates up to 7 syllables per second, and another study that verified that even advanced non-native speakers have increasingly serious problems in hearing and identifying both content and function words in sentences from TV shows spoken at increasingly higher speech speeds, as they missed 21.2% of the words spoken at 6 syllables per second and 32.7% at 7 syllables per second. Using the results of the previous studies, we determine a Word Recognition Ratio of 85%, which indicates how much of the speech in a movie advanced non-native speakers might be able to recognize, given opportunities for repeated listening, and given that there is no unknown vocabulary. We find that this is far from an adequate ratio for comprehension. We caution that measuring listening ability is not the same as measuring listening comprehension. We comment on some
factors affecting listening ability and comprehension, and remark on some ideas for improving listening skills.