Influences of Pause Duration on Impressions of English Speech: Comparison Between Native and Non-Native Speakers

劉, 詩濛

出版情報：九州大学, 2022, 博士（芸術工学）, 課程博士
バージョン：
権利関係：

KYUSHU UNIVERSITY

Graduate School of Design

Department of Human Science

Human Science International Course – Doctoral Program

Influences of Pause Duration on Impressions of English Speech: Comparison Between Native and Non-Native Speakers

ポーズ時間長が英語音声の印象にもたらす影響：母語話者と非母語話者の比較

A DISSERTATION

SUBMITTED TO THE DEPARTMENT OF HUMAN SCIENCE

AND THE COMMITTEE ON GRADUATE STUDIES

OF KYUSHU UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Shimeng Liu

3DS18009K

June, 30, 2022

Principal Advisor:

Dr. Gerard B. Remijn

Acknowledgements

First of all, I would like to extend my sincere gratitude to Prof. Dr. Gerard B. Remijn and Prof. Emeritus Yoshitaka Nakajima for their kindnesses continuous support of my doctoral study and research professionally and patiently, for encouragement and professional knowledge supporting me in finishing this dissertation. High tribute shall be paid to Prof. Dr. Kazuo Ueda. He gave me useful and detailed suggestions about my research.

Special thanks should go to Prof. Dr. Mark Elliot and Prof. Dr. Lihan Chen for supporting this international project continuously. I would also like to thank the students in their laboratory, especially Dr. Sophia Arndt, Xiao Lei, and Yiru Liu for recruting participants and technical help. Furthermore, I appreciated all the students help and efforts from the Professor Nakajima Laboratory (October 2016 – March 2020), from the Professor Remijn Laboratory, and from the Professor Ueda Laboratory, especially for Takuya Kishida for coding the program and technical assistance, Maki Kakizoe, and Mingji Lin for their support in data collection and analysis, and other students for their friendship and to help me with their valuable data for my listening experiments.

Moreover, the most important thanks for my family, especially to my father Yukai Liu, my mother Yan Hu, for their loves and supports in each day. I am also deeply indebted to my friends, Yukai Liu, Jingya Li, and Xu Chen for the generous help for my work.

Last but not the least, this thesis is in loving memory of Yuko Yamashita, who inspired me for this study.

Abstract

The purpose of this study was to investigate how the subjective impression of English speech would change when pause duration at punctuation marks was varied. Two listening experiments were performed in which English speech segments were rated on a variety of evaluation items by both native-English speakers and non-native speakers (native-Chinese speakers and native-Japanese speakers). The ratings were then subjected to factor analysis. In the first experiment, the pauses in three segments were made into the same durations, from 0.075 to 4.8 seconds (s). Participants rated the segments on 23 evaluation items on a rating scale from 1 to 10. A varimax rotation after PCA (principal component analysis) led to two factors that were related to speech style. These two factors could be interpreted as representing speech naturalness and speech rate. Speech segments with a pause duration of 0.6 s received the highest naturalness evaluation, while speech rate perceptually decreased as the physical pause duration increased, without any changes in utterance segments. In the second experiment, a full-factorial design of pause durations (0.15, 0.3, 0.6, 1.2, and 2.4 s) within and between sentences, i.e., for commas and for periods, was implemented in two speech segments. The original speech segments and speech segments without any pauses were also included for control conditions. From ratings on 12 evaluation items, similar to Experiment 1, two factors representing speech naturalness and speech rate were obtained. The results showed again that the perceived speech rate decreased with an increase only in pause duration. As for speech naturalness, the highest evaluations occurred when pause durations were 0.6 s within sentences, and either 0.6 or 1.2 s between sentences. This indicates that fixing all pause durations to 0.6 s is a practical way to train non-native speakers to make their spoken English more natural.

The results may also be of practical use in the development of artificial speech technology, regarding both speech generation and recognition.

Research Output

The contents of the following Chapters of this thesis have been presented, partially or in full, in the following conference proceedings or academic peer-reviewed journal.

1. Chapter 2 and Chapter 3 were presented in an oral session of the 35th Annual Meeting of the International Society for Psychophysics, Fechner Day 2019, Antalya, Turkey, 30 October–2 November 2017.

   Liu, S., Nakajima, Y., Elliott, M.A., Chen, L., Remijn, G.B., Arndt, S., Pang, Z. (2019). Pause Duration Influences Impressions of English Speech Style Rated by Native and Non-native Speakers. In du Bois, N., Arndt, S., Özsoy, E.V., Bayraktar, S., Gülbetekin, E., Elliot, M.A. (Ed.), Conference Proceedings, 35th Annual Meeting of the International Society for Psychophysics, Fechner Day 2019 (pp. 14).

2. Chapter 2 and Chapter 3 were published in: Frontiers of Psychology.
   Liu, S., Nakajima, Y., Chen, L., Arndt, S., Kakizoe, M., Elliott, M. A., and Remijn, G. B. (2022). How Pause Duration Influences Impressions of English Speech: Comparison Between Native and Non-native Speakers. Frontiers of Psychology. 13:778018. doi: 10.3389/fpsyg.2022.778018

Table of Contents

List of Figures

List of Tables

List of Abbreviations

CET College English Test

dBA Decibel-A weighting

EFL English as a foreign language

Hz Hertz

IELTS International English Language Testing System

kHz Kilohertz

ms Millisecond

s Second

SODA substitutions, omissions, distortions, and additions

TFS Temporal fine structure

TOEFL-iBT Test of English as a Foreign Language-Internet-based Test

TOEFL-ITP Test of English as a Foreign Language-Institutional Testing Program

TOEIC Test of English for International Communication

VOT Voice onset time

Chapter 1 - Introduction: Speaking in Public and Pausology

Research into speech pause production and perception has a long tradition, with Tosi (1965) the first to introduce the word *pausology* in a study on speech and music. This was then adopted by O'Connell and Kowal (1983) who investigated silent pauses when reading aloud. O'Connell and Kowal identified various factors that could influence the voluntary use of pauses in speech, such as the speaker's need to breathe, his/her emotional condition, the syntactic complexity of the text, the availability of lexical items, emphasis, and many others (see also Todd, 1985; Lucas, 2015; Barry, 2017). Pauses also play a role as turn-taking in communicative interactions (Sacks et al., 1974; Taneichi, 2014).

Mandatory pausing points are made at punctuation marks, which are used to give meaning and clarity to a sentence, or to separate phrases (Straus et al., 2014). Essentially, their main function is to group speech elements into units (Grosjean et al., 1979; Goldman-Eisler, 1972; Oliveira, 2002). The most commonly used punctuation marks for pausing are periods and commas. In the present doctoral thesis, the use of pauses and their duration in English speech is investigated. In this chapter, first an overview is given on the delivery of public speeches (Chapter 1.1), elements of good delivery (Chapter 1.2), and the most common types of silent parts and pauses in speech (Chapter 1.3). This is followed by an overview of punctuation marks in three different languages (Chapter 1.3.3) and previous research on the perception and production of pauses at punctuations marks (Chapter 1.4).

1.1 Speech delivery

Speech delivery is way of verbal and nonverbal communication, for which a speaker utilizes his or her body language to make their verbal presentation more vivid (Wrench et

al., 2016). The main purpose is to deliver a message, that is, to convey thoughts or ideas to the listeners in an intelligible way. Basically speaking, there are four methods of speech delivery (Lucas, 2015): reading verbatim from a manuscript, reciting a memorized text, speaking impromptu, and speaking extemporaneously, i.e., delivering a carefully prepared and rehearsed speech that is presented from brief notes or outline (Lucas, 2015). The first type of speech delivery consists of reading a fully scripted speech (Wrench et al., 2016). In certain situations, such as in the case of a pope's religious proclamation, the presentation of a scientific report in a professional conference, or a president's rhetorical speech, a speech must be delivered word for word, based on a well-prepared manuscript. In this way, a speaker should work on speech delivery skills that show that the speaker is talking to his or her audience rather than just reading to them. The second type of speech, the memorized speech, is the rote recitation of a written message that the speaker has committed to memory (Wrench et al., 2016). Memorized speech is often used for common occasions or situations, such as for toasts, as congratulatory remarks, prize or position acceptance, or introductions. Here a short memorized speech is usually sufficient, but also these speeches ask from the speaker to focus on communicating with the audience instead of just reciting the words (Lucas, 2015). Memorized speeches, however, are not only used for common occasions. Throughout history, many examples of extraordinary speeches are known in which the orators delivered the whole speech from their memory, and these speeches even had the power to be remembered as an important historical event (e.g., Martin Luther King's speech). The third speech type, the impromptu speech, is a speech delivered with little or no immediate preparation. Typically, a speaker will be asked to give an impromptu speech in the following situations: in a class discussion, a business meeting, or a committee report. A perfect speech is actually not expected by the audience,

but the point should be passed to the audience clearly via the following steps: the speaker must first state the point that he or she is answering, state his/her own point, put forward the appropriated evidence to support the point, and finally summarize the point. To become an outstanding impromptu speaker, lots of practice should be done (Lucas, 2015). Comparing to the impromptu speech, the extemporaneous speech is another kind of speech which requires careful preparation and practice. When presenting this kind of speech, a speaker may use a brief set of notes or a speaking outline to assist his or her memory. An extemporaneous speech has many advantages over other methods of speech. It provides better management over thought and language compared to an impromptu speech, and gives more spontaneity and directness compared to the speech from memory or manuscript. Besides, it adapts to more speech occasions, and encourages the conversational quality through which the audience could provide good interaction with the speaker (Lucas, 2015). Even though the aforementioned methods of speech delivery are a little bit different from each other, a good delivery is asked for all speech types. Good speeches must be delivered conveying the speaker's idea clearly, interestingly, and without distracting the audience (Lucas, 2015).

1.2 Elements of good delivery

Good delivery is a process of presenting a clear, coherent message in an interesting way (Wrench et al., 2016). Another obviously important point is that the speech should be delivered in an understandable way. Understanding the message of the speech concerns not only content, but also nonverbal aspects. There are six elements of good delivery: conversational style, conversational quality, eye contact, vocalics, physical manipulation, and variety (Wrench et al., 2016).

Conversational style

Conversational style is a speaker's ability to sound expressive and to be perceived by the audience as natural (Wrench et al., 2016), while **conversational quality** is the ability that makes the rehearsed speech sound spontaneous (Lucas, 2015).

Eye contact

Eye contact is a speaker's ability to maintain visual contact with the audience (Wrench et al., 2016). It also can help speakers to establish a good relationship with an audience (O'Hair et al., 2001), and convey a wide range of emotions, e.g., sadness, compassion, concern, anger, and joy (Koch, 2010).

Vocalics

Vocalics is the subfield of nonverbal communication that examines how a speaker uses his or her voices to communicate orally, and it is also known as paralanguage (Wrench et al., 2016). Speakers should deliver the speech loudly enough for all audience members (even those sitting in the back of a room), and enunciate clearly enough to be understood by all audience members (even those who may have a hearing impairment or who may be English-language learners). Wrench et al. (2016) gave the following instructions of good vocalic technique for all speakers:

(1) Face the audience with chin up.

(2) Keep the eyes away from the notecards.

(3) Set the voice at a moderate speed.

These methods could help speakers with their vocalics during the delivery of speech.

Effective use of vocalics also means that a speaker should make use of appropriate pitch, pauses, vocal variety, and correct pronunciation (Wrench et al., 2016).

Volume refers to the loudness of a speaker's voice. As mentioned, public speakers need to speak loudly enough to be heard by everyone in the audience. In addition, volume is often needed to overcome ambient noise, e.g., the hum of an air conditioner, or the noise of traffic passing by. Besides, a speaker can also use volume strategically to emphasize the important information in speech (Wrench et al., 2016).

Rate is the speed at which a person speaks. To keep the speech delivery interesting, speech rate should vary. Different speech rates can present different emotions (Wrench et al., 2016). A rapid, lively rate can communicate such meanings as enthusiasm, urgency, or humor; while a slower, moderated rate can convey respect, seriousness, or careful reasoning. By varying rate within a single speech, a speaker can emphasize the main points and keep the audience interested.

Pitch refers to the perceived height of a speaker's voice or note. Changes in the pitch or tone of a speaker's voice are known as inflections (Lucas, 2015). A speaker can use pitch inflections to make the delivery more interesting and emphatic (Wrench et al., 2016).

Vocal variety refers to changes in the vocalics, e.g. volume, pitch, rate, and pauses that give the voice variety and expressiveness (Lucas, 2015). Vocal variety should flow naturally from a speaker's wish to speak with expression. In that way, it will animate the speech and invite listeners to understand the topic easily (Wrench et al., 2016).

Since pronunciation (articulation) and pauses are directly related to the present research topic, these two aspects of vocalics are discussed in more details separately [pronunciation (articulation) in 1.2.1, pauses in 1.2.2].

Effective physical manipulation

Besides using vocalics aspects effectively, another element that makes a good speech delivery is physical manipulation, which means using body language to emphasize meanings or convey meanings during a speech (Wrench et al., 2016). There are four basic aspects of physical manipulation: posture, body movement, facial expressions, and dress. These aspects add up to the overall physical dimension of speech delivery, which is called self-presentation.

Posture is quite important during the speech (Wrench et al., 2016). When communicating to the audience, as long as a speaker stands up straight, even without saying a word, it can make the audience convinced that the speaker is holding a position of power and taking his or her position seriously. Body movements include steps and gestures (Wrench et al., 2016). Movements at transition points help the audience not only to focus attention on the transition from one idea to the next, but also help the speaker to increase nonverbal immediacy by getting closer to different segments of an audience. Using appropriate gestures can be an efficient way to suggest emphasis, enthusiasm, or other personal connection with a speech topic. However, overusing movements and gestures could also obscure meaning.

Facial expressions can convey much information, so a speaker must be acutely aware of what his or her face looks like during the speech. No facial expression and over animated facial expressions are the two extreme things that a speaker should avoid. Facial expression can be used strategically to enhance meaning, and should be consistent with the contents that the speaker is delivering (Wrench et al., 2016).

Dress is still a very important part of how audience will perceive a speaker during the speech (Wrench et al., 2016). One general rule for a speaker to determine his or her dress

is the “step-above rule,” which states that a speaker should dress one step above the audience. Another general rule for dressing is to avoid distractions in the speaker’s appearance.

Self-presentation, referring to poise or stage presence, is determined by how a speaker looks, how a speaker stands, how a speaker walks to the lectern, and how a speaker uses his or her voice and gestures. Self-presentation should support a speaker’s credibility and improve the likelihood that the audience will listen with interest (Wrench et al., 2016).

Variety

The last element of good speech delivery is variety (Wrench et al., 2016). During the speech, a speaker should include a variety of different nonverbal components. In addition, a speaker should make sure that his or her face, body, and words are all working in conjunction with each other to support his or her message (Wrench et al., 2016).

1.2.1 Pronunciation

Pronunciation concerns the conventional patterns of speech used to form a word. Understandable pronunciation is one of the basic requirements of language learners’ competence and it is also one of the most important features of language instruction (Gilakjani, 2012). In the speech delivery, word pronunciation is important for two reasons: First, mispronouncing a word harms the credibility as a speaker; and second, mispronouncing a word they are unfamiliar with can confuse and even misinform the audience (Wrench et al., 2016).

One important aspect of pronunciation is articulation, and it is the ability to clearly pronounce each of a succession of syllables used to make up a word. In American English,

the acronym SODA (substitutions, omissions, distortions, and additions) is utilized to help people learn how to articulate more effectively (Wrench et al., 2016), and it essentially constitutes a categorization of articulatory errors (Dawadee et al, 2017).

Substitutions: One or more consonants or vowels are substituted for another consonant; e.g., wabbit for rabbit, tow for cow, mouf for mouth.

Omissions: Certain sounds are not produced within a word - entire syllables or classes of sounds may be deleted; e.g., fi' for fish, 'at for cat, pro'lly for probably.

Distortions: Sounds are changed slightly with nasal or slurring sounds so that the intended sound may be recognized but sounds "wrong," or may not sound like any sound in the language. Pencil sounds like mencil, second sounds like slecond.

Additions: An extra sound or sounds are added to the intended word; e.g., anyways for anyway, athaletic for athletic, buhlack for black.

1.2.2 Pauses

Another important factor for delivering a good speech is dealing with punctuation and pauses. Pauses play an important role in speech delivery, and learning when and how long to pause is a very basic thing for beginning speakers (Lucas, 2015). Appropriate pausing can signal the end of a thought unit, give time for listeners to think or even lend a dramatic impact to a statement. Lucas (2015) only gave the direction that speakers should pause at the end of thought units instead of in the middle, and he attributed timing control more as a matter of sense or partly a matter of experience. However, Lucas (2015) did not give any precise instructions on how long to pause.

Both Lucas (2015) and Barry (2017) pointed out that utilizing digital devices (i.e., a smart phone, a tablet) could be useful to record the speech performances in order to

practice speech delivery in self-training. With regard to pausing during the delivery of speeches, Barry's (2017) instructions for practicing speech with the manuscript displayed on a telephone are as follows:

(1) Look at the phone for 2 s in silence.

(2) Then look down at the manuscript and quickly bear several words in mind.

(3) Next, look up from the phone, head up to the imaginary audience, pause again in silence for completely 2 s, and deliver the words that picked up before.

(4) Again, pause for another 2 s completely, in the meanwhile look at the device.

(5) Then look down to the manuscript and memorize several new words, and deliver them with your head up to the audience.

(6) Repeat these steps, varying the number of words that are kept in mind and deliver, until the speech reaches at least 2 minutes in duration. Finally, extend the duration of the delivery, until a whole delivery can be accomplished.

According to Barry (2017), when the speaker becomes familiar with this way, he or she can vary the number of words and seconds. Even though Barry (2017) did not give the exact suitable pause duration for pause between ideas, he emphasized the importance to pause at commas and periods. Barry (2017) also pointed out that the speaker's job is to let the audience think rather than talk, and the only time for thinking is during pauses.

1.3 Silent parts and pauses in speech

Speech contains many silent parts that need to be articulated in order to be understood by the listener. There are many kinds of silences in speech, based on duration and position. From short to long duration, the most common ones are articulatory silences, such as voice-onset time (VOT) in speech syllables, word segmentation between words, and

punctuation between phrases. These three silences affect how people perceive speech in different dimensions.

1.3.1 Articulatory silent parts, including voice onset time (VOT)

Figure 1.1 shows examples of waveforms of (a) VOT (voice on set time), (b) word segmentation, and (c) punctuation. These are examples of silent parts or pauses in speech. Many other short silences exist in speech that appear within or in between neighboring stop consonants, fricatives, and affricates (i.e., obstruents). Silent parts before stop consonants, as the silence between /aɪ/ and /k/ in 'icon', or before the /t/ of "top" in the phrase "at the top", are necessary to perceive the words correctly. The duration and the place of the silences can influence how utterances are perceived. Indeed, research has shown that a silence is necessary before a stop, in order to perceive a stop, for example, as the second element in a fricative-stop-vowel syllable. Similarly, a silence is necessary after a stop in syllable-final position, in order to perceive the stop when a second syllable is added. Interestingly, even if there are no clear spectral cues, a stop or affricate can be perceived if a silence is added at an appropriate position (e.g., Dorman et al., 1979; Liberman, 1996). Even short thus silences can change the perception of speech. The speaker can even actively manipulate the duration of these silences to express a certain affect or to clearly articulate.

The shortest type of silences that affect the perception of utterances is voice onset time, or VOT. Voice onset time (VOT) is the time interval between the burst that marks the release of the stop closure and the onset of quasi-periodicity which reflects laryngeal vibration. An example of VOT is shown in **Figure 1.1a**. The reason why "pa" is shown as an example of VOT is that "pa" is part of the word "passion", and it is from Speech

Segment 3 used throughout this thesis. Moreover, it is also a typical example of VOT (Lisker & Abramson, 1964). VOT plays an unambiguously distinctive role in the case of perceiving consonants in most languages (Lisker & Abramson, 1964). The VOT values of voiceless stops are higher than those for the voiced stops overall (Lisker & Abramson, 1964). For the present purpose, VOT is thus a typical example in which a silence changes the perception of a speech segment; here, longer VOT silences are related to the perception of voiceless stops. The relationship between VOT values and voice or voiceless stops differs among language groups (Lisker & Abramson, 1964; Cho & Ladefoged, 1999; Keating, 1984). In their experiment, Lisker and Abramson, used two methods to classify English stop consonants /k/, /g/, /t/, /d/, /p/, and /b/. In the first method, the consonants were labeled as velar stops /k/, and /g/, alveolar stops /t/ and /d/, and bilabial stops /p/ and b/. In the second method, the consonants were separated into voiced stops /g/, /d/, and /b/ and voiceless stops /k/, /t/, and /p/, according to the presence or absence of glottal buzz. As a result, for both voiced and voiceless stops, velar stops (voiced: from -108 ms to 21 ms, voiceless: from 24 ms to 80 ms) tended to get higher mean VOT values than alveolar stops (voiced: from -110 ms to 5 ms, voiceless: from 8 ms to 70 ms) and bilabial stops (voiced: from -138 ms to 1 ms, voiceless: from 2 ms to 58 ms). VOT values also differ significantly between age groups (e.g., Kang, 2014) and gender groups (e.g., Li, 2013), and may depend on country region (e.g., Takada, 2008). Taken together, research has led to the conclusion that VOT has a considerable influence on distinguishing consonants among different words, even in different languages.

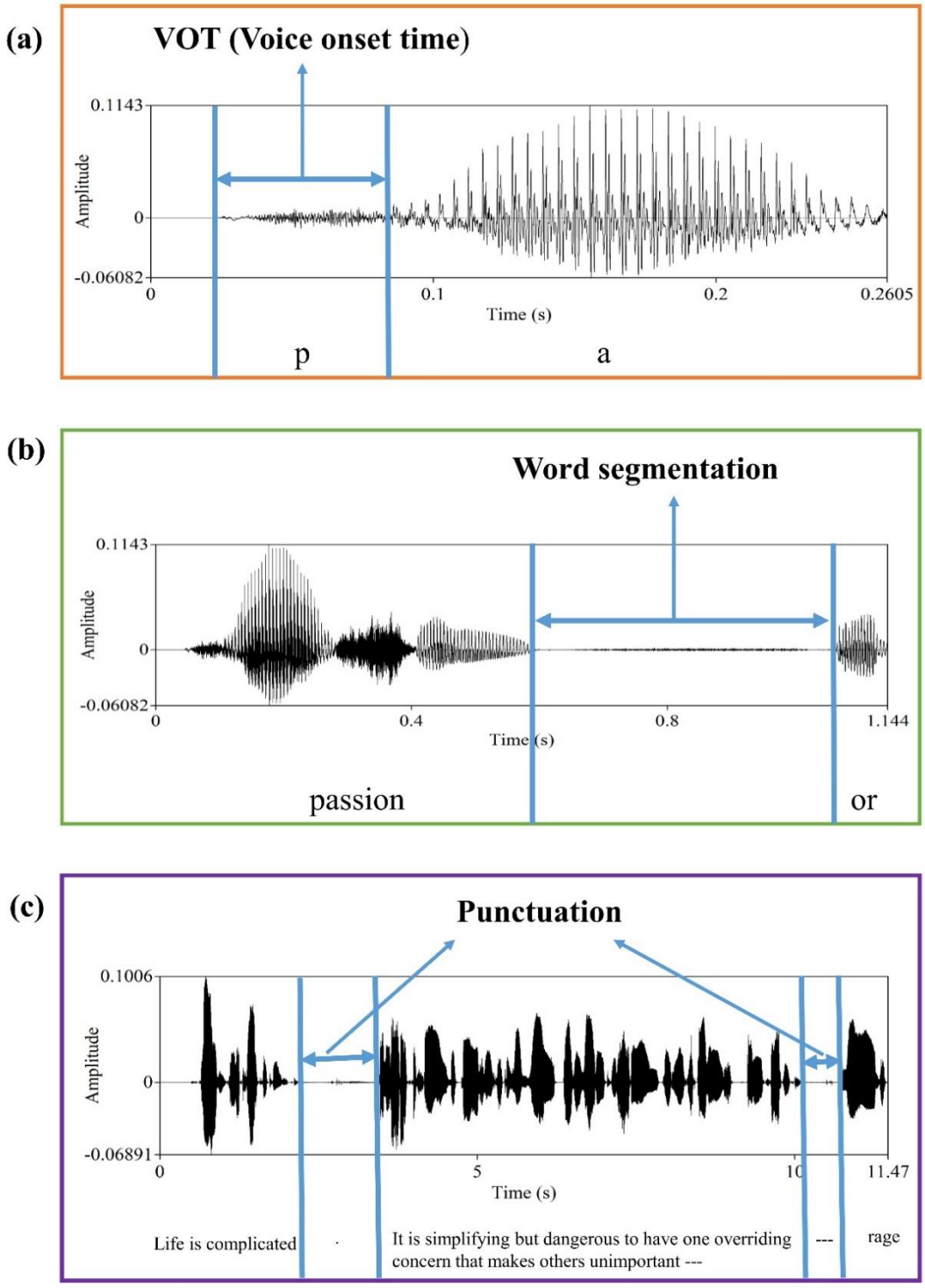


Figure 1.1 Examples of waveforms of (a) VOT (voice on set time), (b) word segmentation, and (c) punctuation. (Waveform source: Sentences from original speech of Speech Segment 3 in Experiments 1 and 2.)

1.3.2 Word segmentation

The second type of silence or pause in speech concerns word segmentation (**Figure 1.1b**). Even when there is no silent part in between words, listeners can perceptually insert a silent part. This is called word segmentation. However, word segmentation may have different meanings in different languages. Words are generally considered to be basic meaningful units of a language (Liu et al., 2013). For example, in alphabetic language writing systems, such as English, unambiguous markers, e.g., interword spaces, were used to segment sentences into individual words. However, in language systems that use characters, such as Mandarin Chinese, the character is the most basic and meaningful unit of information. There are no unambiguous markers between written words, thus in written Mandarin Chinese, there are no explicit cues to tell readers where a word begins or ends in a serial string of characters (Liu et al., 2013).

As regards word segmentation in English, a study by Malt and Seamon (1978) showed that reading speed was significantly reduced (at least by one syllable / second or from 1 syllables/s to 1.7 syllables/s) when the spaces between words were filled with black, English-letter-like fillers for English conditions. Thus, inappropriate word segmentation reduced reading speed and increased reading errors. Bai et al. (2008) investigated the use of different reading speeds by Chinese native speakers, using four conditions: normal unspaced text, text with spaces between words, text with spaces between characters that yielded nonwords, and finally text with spaces between every Chinese character. From the results it could be seen that there was a decrease of sentence reading when a space was inserted between characters (from 239 characters / minute to 220 characters / minute). However, no influence was observed when spaces were inserted between words (239

characters / minute), compared with no spaces inserted (239 characters / minute). However, a lower reading speed (213 characters / minute) was obtained when a space was inserted between non-words that were set as the target items. The results proved that, rather than individual characters, words units are the unit of information in Mandarin Chinese. Moreover, the results imply that Chinese readers have a flexible comprehension of word segmentation processing. Therefore, word segmentation certainly has an effect on the reading speed and comprehension of the speech content.

1.3.3 Punctuation

The third type of silence in speech are the easiest to perceive, and are made a punctuation marks (**Figure 1.1c**). Pauses at punctuation are the longest silences in speech, and they are used to separate phrases and sentences (Goldman-Eisler, 1972; Grosjean et et al., 1979; Oliveira, 2002). They are also used for physical reasons, e.g., breathing (O'Connell & Kowal, 1983). Punctuation refers to the part where the sentence breaks and the part where the sentence is cut off when reading the sentence. A symbol placed with a certain rule within or between sentences is called a punctuation mark. The most typical ones are commas and periods, and they are used almost every day in many languages. An overview of punctuation marks in English, Mandarin Chinese, and Japanese is given below. Punctuation marks in these three languages were picked up because both in Experiment 1 (described in Chapter 2), and in Experiment 2 (described in Chapter 3), speakers from three different language backgrounds, i.e., English, Chinese and Japanese speakers, were invited to join the experiments. Another reason was that they each represent a different type of language: English is a stress-based language, Chinese a

syllable-based language, and Japanese a mora-timed language.

Punctuation in English– a stress-based language

English is one of the most widely used languages in the world. In modern English, there are 13 types of punctuation marks according to the instructional document "The Blue Book of Grammar and Punctuation" regarding English grammar and punctuation (Straus et al., 2014). The punctuation marks are the following:

(1) Period **.**

A symbol used at the end of a completed sentence to represent a statement or after an abbreviation.

(2) Comma **,**

This is an in-sentence symbol that indicates a short pause (i.e., for word, clause segmentation).

(3) Semicolon **;**

Similar to a comma, it is an in-text symbol indicating a longer pause (i.e., association of related phrases, comma-related word segmentation).

(4) Colon **:**

This is a symbol used to mean “that means” and “I mean this way”.

(5) Question mark **?**

This is a symbol used after a direct question.

(6) Parentheses **( )** or **[ ]**

These symbols are parentheses and square brackets.

・Parentheses: Symbols used to present ancillary information.

・Square brackets: Symbols used in limited conditions, used exclusively for quoted data.

(7) Apostrophe ’

This is a symbol used for affiliation or abbreviation.

(8) Hyphen -

Unlike the dash, it is a symbol that mainly combines portions of multiple words.

(9) Dash —

Longer than a hyphen, it is a symbol used when adding emphasis, placing a break, or suddenly changing thoughts.

(10) Abbreviation code …

This symbol is mainly used when omitting words, phrases, lines, paragraphs or longer quoted clauses. It is also used for hesitation and change of mood.

(11) Quotation marks “ ” or ‘ ’

This is a symbol mainly used when quoting content (i.e., a story, literature). Single quotes are used inside double quotes.

(12) Exclamation mark !

This is a symbol used to express emotions, emphasis or surprise (not used when writing formal sentences).

(13) Diagonal line /

This is a symbol used when entering a date or separating a line of poetry. (This symbol is also used in cases as he/she, and/or, increase/decrease, etc.)

Among the aforementioned punctuation marks, commas and periods are the most commonly used marks in English. Research about pause duration at punctuation marks in

English will be introduced in Chapter 1.4.

Punctuation in Chinese – a syllable-based language

While English is the most representative of alphabetic languages, Chinese is the most representative of syllable-based languages which have roughly equal syllable durations (Pike, 1945). In Chinese, people distinguish the meaning of the characters or word phrases by their tone. In Mandarin Chinese, there are five tones (i.e., four tones and one neutral tone) in total. The four tones are called “first tone”, “second tone”, “third tone”, and “fourth tone”, respectively. The first tone is a high-level flat tone. The second tone is a rising tone. The third tone is a low tone and demonstrated as having a rise in pitch after the low fall. The fourth tone is a falling tone. Finally, the neutral tone is sometimes thought of as having a lack of tone direction and has a weaker pronunciation. An example to show that the meaning of the character changes with its pronunciation is the pronunciation of the Mandarin Chinese character “ma”. According to the tone, from the first to the fourth tone, respectively, the meaning could vary from “mother”, “hemp”, “horse”, to “scold”. The neutral tone of “ma” indicates that a question is made (Lin et al., 2021).

Chinese characters have been handed down for thousands of years. However, punctuation marks in Chinese do not have such a long history. Since there were no punctuation marks in ancient Chinese, phrases were separated by common sense, based on context. International punctuation was first implemented in Chinese in 1920, and was used for the common Chinese language at that time. The current used punctuation marks for modern Chinese are a revised version issued in 2011 based on the national standard “general punctuation rules”, which was issued by the Chinese government in March, 1990.

Due to its content, modern Chinese possesses a total of 17 types of punctuation marks, for the present purpose of the thesis, only punctuation marks that cut texts into portions are listed (Emphasized symbol and underline are not listed). According to the General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China & Standardization Administration [GAQSIQ & SA] (2011), they are:

(1) Period **.**

It is a symbol used at the end of a sentence to express a statement.

(2) Question mark **?**

It is a symbol used at the end of a sentence to express a question.

(3) Exclamation mark **!**

It is a symbol used at the end of a sentence to express a strong mood (i.e., praise, joy, anger, interjection, surprise, sorrow, appreciation, excitement).

(4) Comma **,**

This is a symbol used within a sentence to represent a median pause within a sentence or paragraph.

(5) Comma dot **、**

This is a symbol used within a sentence to represent a relatively shorter pause between parallel words in a paragraph, or a pause after a number which represents a multi-level order.

(6) Semicolon **;**

This is an in-sentence symbol used to represent a pause between parallel clauses in a compound sentence, or a pause between the lowest layer clauses of a non-parallel multiple duplicate sentence.

(7) Colon **:**

This is a symbol used within a sentence to present the following sentence or summarize the aforementioned sentences.

(8) Quotation marks **“ ”** or **‘ ’**

This is a symbol used to represent a directly quoted part or a part that should be pointed out in a paragraph.

(9) Parentheses **( )** or **[ ]** or { }

These are symbols used to represent an annotation part, supplementary explanation part, or other special meaningful sentences in a paragraph.

(10) Dash ——

A symbol used to present a comment, supplementary explanation, or change in voice or meaning in a certain part of a paragraph.

(11) Abbreviation code **……**

This is a symbol used to indicate the omission of certain content in a paragraph or the interruption of meaning.

(12) Hyphen **-** or — (It is half long as dash) or ～ ”

This is a symbol used to present the connection to a certain related part (“—” is as half long as a dash).

(13) Bullets ·

This is a symbol used to present the boundary between certain related parts.

(14) Guillemets **<< >>** or **< >**

This is the symbol used to present the name of the literature work that appears in the paragraph.

(15) Separatrix mark /

A symbol used to indicate a poem phrase, time signature, or a delimiter for a related character (This mark is not existing in Japanese).

In Chinese, not only commas and periods, but also the following punctuation marks are commonly used as in English question marks, exclamation marks, semicolons, parentheses, and dashes. There are also some punctuation marks that do not exist in English (i.e., comma dots, and guillements).

Punctuation in Japanese – a mora-timed language

Japanese is representative of a mora-timed language, which is somehow between a tonal language and a stress language (Ladefoged & Johnson, 2006). The mora is a unit of timing, and each mora has approximately the same duration of 120 ms (Kohno, 1998; Ladefoged & Johnson, 2015). It also can be seen as the unit of phonological distance, since a long syllable consists of two morae, and a short syllable consists of one mora in Japanese (McCawley, 1968; 1978). Overlap exists between the mora and the syllable, and in this case, each mora corresponds to a syllable. However, the mora and the syllable do not always overlap. The reason is that some morae cannot form a syllable on their own (Kubozono, 2015).

There are 10 types of punctuation marks in Japanese (Croes, & Dexter, 2016):

(1) Comma **,** or 、

This symbol is used in many contexts, principally for marking off separate elements within a sentence. It is called “tōten” too in Japanese.

(2) Period **.** or 。

This symbol is often used to separate consecutive sentences. It is called “kuten” in Japanese.

(3) Colon **:**

This is a symbol used within a sentence to inform the reader that what follows proves, clarifies, explains, or simply enumerates elements of what is referred to before. It is called “koron” in Japanese.

(4) Exclamation mark **!**

This is a symbol usually used after an interjection or exclamation to indicate strong feelings or high volume, and generally marks the end of a sentence. It is called “kantanfu” in Japanese.

(5) Question mark **?**

This is a symbol commonly used in comics and creative writing at the end of a sentence to present a question mood. It is called “gimonfu” in Japanese.

(6) Ellipsis **…**

This is a symbol used at the end of a sentence, and indicates an intentional omission or abbreviation, a pause in speech, an unfinished thought or a trailing off into silence (aposiopesis). It is called “tensen” in Japanese.

(7) Interpunct ・

This symbol is a small dot used for inter-word separation. It is called “nakaguro” in Japanese.

(8) Quotation marks 「 」 or 『 』

Single quotation m/arks are used for indicating quotes, while double quotes are a lot less common than single quotes, and they are used inside single quotes when

quoting text within quoted text. They are called “kagikakko” and “nijūkagikakko” in Japanese.

(9) Wave dash 〜

This symbol is used to show a range of something, draw out and change the pitch of a vowel sound (そうだね〜 “sou da ne”, with elongated “e” of “ne”), show where something is from, and mark subtitles. It is called “namisen” in Japanese.

(10) Parentheses ( ), [ ], 【 】, or ｛ ｝

These symbols are used for annotations (like this) within a sentence. Among these parentheses “ ( ) ” are the most commonly used ones. Square brackets “ [ ] ” do not have a singular use, and they can be used for a wide range: showing emphasis, listing items, or just making brackets themselves stand out more. Lenticular brackets “【 】” are widely used in Japanese dictionaries and other educational resources. Wave brackets “ ｛ ｝ ” are basically used inside normal brackets, like “ ［｛｝］ ” and for mathematical equations. They are called “marukakko”, “kakukakko”, “sumitsukikakko”, and “namikakko”, respectively.

In Japanese language, there are also some punctuation marks used in a similar way as in English (i.e., commas, periods, and exclamation marks). The quotation marks in Japanese have the same meaning as in English, but the writing styles are different. The wave dash is not really similar to the straight dash in English.

Overall, although there are fairly different uses of punctuation marks among these three language groups, commas and periods are common in English, Chinese, and

Japanese to indicates pauses.

1.4 – Research on pausology

Research into speech pause production has shown that pause duration varies with communication style (O'Connell & Kowal, 1983). For example, for story telling in English, a mean pause duration of 0.94 s (SD= 0.23 s) was found for segments with a minimum cut-off in between 0.20 - 0.31 s, including commas and periods. However, in interviews the mean pause duration was 0.53 s (SD= 0.06) [Kowal et al., 1983]. For English and Spanish narratives, the mean pause durations were 0.69 s and 0.73 s, respectively (de Johnson et al.,1979), while in poetry readings in English and German, the longest pause duration was used for punctuated line-ends, with a mean duration of 0.71 s [O'Connell & Kowal (1984)].

As discussed, mandatory pausing points are made at punctuation marks, which are used to give meaning and clarity to a sentence, or to separate phrases (Straus et al., 2014). Essentially, their main function is to group speech elements into units (Goldman-Eisler, 1972; Grosjean et al., 1979; Oliveira, 2002). Data from separate analyses of comma- and period-pause durations showed that pause durations between sentences (i.e., periods) are longer than in between clauses within sentences (e.g., commas; Cruttenden 1986). For example, in oral deliveries of sermons in German, the mean duration for commas was 0.47 s (SD= 0.22), while for periods it was 0.98 s (SD= 0.34) [O'Connell & Kowal, 1986]. Interestingly, the average comma and period durations in four university commencement speeches in English were similar to these durations, i.e., 0.49 s (SD= 0.26) for commas and 1.01 s (SD= 0.40) for periods (Yamashita & Fuyuno, 2015). Finally, public presentations in English showed an average pause duration of 0.38 s (SD= 0.22) within sentences and of 0.98 s (SD= 0.33) between sentences. For another script, pause duration

within and between sentences was 0.45 s (SD= 0.31) and 0.81 s (SD= 0.31), respectively (Yamashita et al., 2019). Taken together, the research on a variety of studies on pausing in speech has shown that the mean physical durations of commas (range: 0.38 s to 0.67 s) and periods (range: 0.81 s to 1.24 s) thus typically have a ratio of 1: 2.

Since English is used as a lingua franca (e.g., Jenkins et al., 2011), non-native speakers far outnumber native speakers (Crystal, 2008). For most beginning speakers of a language that is syntactically very different from their first language, it is a primary issue to learn how and when to pause, and to control pause durations. As Handel (1989) argued in his classic chapter on rhythm perception, to control pause durations is very important in speech communication. In preliminary studies on this topic with learners of English as a second language (L2 learners), recordings were obtained from first-year students in an EFL writing and speaking course at two Japanese universities, who practiced English public presentations (Yamashita et al., 2014), or participated in a speech competition (Liu et al., 2016). Temporal factors in their speech were analyzed, including the number of pauses, their median duration and maximum duration, the standard deviation of the pause duration, the coefficient of variations in pause duration within sentences (e.g., commas), and the coefficient of variations in pause duration between sentences (e.g., periods). In these studies, (Yamashita et al., 2014; Liu et al., 2016), the median pause duration ranged from 0.40 to 0.64 s. The maximum pause duration, however, varied considerably between speakers (1.15 to 4.49 s). The coefficient of variations of pause duration reflected the speaker's proficiency: Participants who had a lower coefficient tended to get a higher evaluation in the speech competition (Liu et al., 2016). In the top-3 speeches with the highest evaluations, among a total of 11 speeches, the pause duration within sentences was 0.59 s, 0.42 s, and 0.67 s, while the pause duration between sentences was 1.24 s,

0.83 s, and 1.13 s, respectively. The pause durations between sentences and within sentences thus also had a ratio of about 1:2 for these proficient L2-leaners. By contrast, for the 3 bottom-ranked L2-learners the ratio varied considerably. Their pause duration within sentences ranged from 0.50 to 0.92 s, while it ranged between sentences from 0.86 to 1.54 s, thus with a ratio in between 1:1.72 to 1:2.5. In a related study, pause insertion patterns of English L2-learners were also investigated from a perspective of multimodal corpora (Fuyuno et al., 2016). The relative cumulative frequencies of the duration of pauses in commas and periods of proficient L2-learners were similar to those of native-English speakers. Furthermore, proficient L2-learners demonstrated a similar pause insertion pattern (Fuyuno et al., 2017). Proficient L2-learners also shared similar pause patterns (i.e., number, duration, and location of pauses) in different speech rates in speech production (Matzinger et al., 2020), and no difference in pause duration and distribution compared to their own languages (Black et al., 1966). Pause duration control thus should have contributed to the quality of L2-learners' speaking performance.

1.5 Key research issues of the present study

The research on L2-learners' use of pauses and that on the voluntary use of pauses by native speakers during public speaking (Lucas, 2015; Barry, 2017) strongly suggests that pause duration affects our general impression of speech. "Voluntary" here means that the speaker uses different pauses at different places to make his/her speech delivery impressive to the audience in public speaking. For example, the speaker can leave a relatively long pause at the end of a thought unit, to allow the audience to think. This way, pauses are used as a rehearsal time for short-term memory, especially for the listeners (Sugito, 1990). Barry (2017) also pointed out that the speaker's job is to let the audience

think rather than talk, and the only time for thinking is during pauses. So far, research on the perception of pauses in speech – rather than their production – has mainly focused on the perceptual underestimation or overestimation of pause durations (Stuckenberg & O'Connell, 1988). As was found in a listening experiment on English and German speech, listeners perceived pause occurrence and pause duration differently, depending on whether the physical pause duration was short (< 0.3 s), medium (> 0.3-< 1.0 s), or long (>1.0 s; Stuckenberg and O'Connell, 1988). Listeners tended to overestimate the physical pause duration when it was short, but tended to underestimate medium and long pause durations. The detection of pauses improved when the physical pause duration was long, but even long pauses were not always detected. As for the number of pauses, listeners perceived more pauses than there actually were. Other research has focused on the automatic detection of pauses in speech with computers (Horii, 1983; Goto et al., 1999; Rosen et al., 2010).

Little is known, however, on (1) how systematic changes in pause duration influence subjective impressions of English speech, and (2) whether favorable impressions occur under a common pause duration, for native and non-native speakers.

In order to investigate these issues, in this thesis, two listening experiments were performed using excerpts from English textbooks (see **Chapter 2.2.2**). In Experiment 1, the main purpose was to investigate how pause duration influences the subjective impression of English speech. Both the comma pause and the period pause at punctuation marks were varied with the same steps in a range from 0.075 to 4.8 s. First, in order to ascertain that the selected segments were typical English speech segments, we analyzed their pause durations and the articulation rates. Following this, native-English speakers and non-native speakers (native-Chinese speakers and native-Japanese speakers) were

asked to evaluate the segments on 23 items based on a rating scale from 1 to 10. Data from three language groups were collected to investigate whether a common favorable pause duration existed regardless of the language background. Before the evaluations were subjected to factor analysis, the distribution of evaluations for each item was observed.

In Experiment 2, in order to further investigate how pause duration within and between sentences influences the subjective impressions of English speech, the comma pause and the period pause at punctuation marks were varied independently using a full-factorial design of pause durations (0.15, 0.3, 0.6, 1.2, and 2.4 s). Native and non-native English speakers were employed in the experiment, and asked to evaluate the segments on 12 items based on a rating scale from 1 to 10. The procedures and methods in Experiment 2 were similar as in Experiment 1.

1.6 Structure of the dissertation

Chapter 2 describes Experiment 1 that investigated how manipulated pause duration at punctuation marks (i.e., commas and periods) influenced subjective impressions of English speech. In this chapter, all research methods that were used have been described, e.g., what kind of manipulated pause durations were used, what kind of speech segments were used, how speech segments were edited, and last, how the stimuli were presented to the listeners. In Chapter 2, Experiment 1 is described in which the inserted pause durations were fixed at 0.075, 0.15, 0.3, 0.6, 1.2, 2.4, and 4.8 s. The duration of each pause was the same for commas and periods. The experiment consists of 3 sub-experiments, one for each of three language groups. Nineteen native-English students or employees, 20 native-Chinese students, and 19 native-Japanese students were employed

as participants to the listening experiments. Participants were asked to rate 23 evaluation items **(see Table 2.1)** based on rating scales from 1 to 10. The rating data were first subjected to a factor analysis following Pett et al. (2003). After this, non-parametric tests, e.g., a Friedman two-way analysis of variance by ranks, a Wilcoxon signed-rank test and Holm-Bonferroni tests as post-hoc tests were performed (Field, 2009).

Chapter 3 describes Experiment 2 that further investigated how manipulated pause duration at punctuation marks (i.e. commas and periods) influenced subjective impressions of English speech for commas and periods separately. In Experiment 2, a full-factorial design of pause duration (0.15, 0.3, 0.6, 1.2, and 2.4 s) was implemented. Participants were asked to rate 13 evaluation items **(see Table 2.1)** based on a rating scale from 1 to 10. The experiment also consists of 3 sub-experiments, one for each of three language groups as in Experiment 1. In these experiments, almost the same procedures as described in Chapter 2 were used, and the analysis methods were the same as well.

Chapter 4 is for the general discussion and conclusions. In this chapter, we summarized the findings and provided discussions in this field according to the results of Chapters 2 and 3. Limitations and potential developments of the present study were argued.

Chapter 2 – Experiment 1

2.1 Introduction

A series of studies comparing the use of pauses between native and non-native English speakers showed that proficient EFL learners (English as a foreign language) shared a similar pause pattern (i.e., pause duration) with native-English speakers: Physical period-pause duration in spontaneous speech was approximately twice as long as that of comma-pause duration (O'Connell & Kowal, 1986; Yamashita & Fuyuno, 2015; Liu et al., 2016; Yamashita et al., 2019). Similarities of relative cumulative frequencies of pause duration, pause insertion pattern, number and location of pauses were observed in studies by Fuyuno et al. (2016, 2017) and Matzinger et al. (2010). On the other hand, the ratio of pause duration within sentences and between sentences from non-proficient EFL learners varied considerably (Liu et al., 2016). It is therefore possible that the use of pause duration in English speech by EFL learners can improve with practice, and this might have a favorable influence on the perceived quality and intelligibility of the English speech. So far, however, little is known about how pause duration affects the perception of speech. The primary purpose of this experiment was therefore to investigate whether and how manipulated pause duration at punctuation marks influenced the subjective impression of English speech.

The second purpose of Experiment 1 was to examine whether favorable impressions occur under a common pause duration, for native and non-native speakers. Kowal and O'Connell (1983) examined the commonalities in the use of time, (i.e., speech and articulation rates, physical pause duration, phrase length, and percentage of pause duration/total duration) in five languages (English, Finnish, French, German, and

Spanish) from a corpora of spontaneous speech. Their results suggested the possibility of a language universal for the use of time. However, we still do not know whether commonalities exist from a perspective of speech perception, for example, whether listeners from different language backgrounds prefer a similar pause duration. To investigate this issue, in Experiment 1 both native and non-native speakers of English, i.e., Chinese and Japanese speakers, were asked to participate.

2.2 Method

Experiment 1 consisted of a listening experiment in which the pause durations in three short English speech segments were varied together into the same 7 steps: 0.075, 0.15, 0.3, 0.6, 1.2, 2.4, and 4.8 s. This range included a pause duration (0.075 s) that was shorter than 0.10 s, which is considered as a minimum psychologically functional duration in reading (Hieke et al., 1983). Although Oehmen et al. (2010) utilized 0.01 s as a threshold for manual segmentation in speech, it has been shown that silent intervals of 0.10 s can appear in speech not as pauses, but as silent intervals preceding stop consonants (Suen & Beddoes, 1974). In a study of silences in turn-taking from the view of conversational corpora, Heldner and Edlund (2010) used 0.18 s as the smallest pause duration to minimize the risk of confusing stop closures with pauses. Goldman-Eisler (1968) even suggested a cut-off point of 0.25 s as a threshold to separate hesitation pauses and phonetic stops. More importantly, as described above, previous research on comma- and period-pause duration show that they physically are in a range of about 0.3 – 1.0 s or longer (Liu et al., 2016; Yamashita et al., 2019). The longest pause duration (4.8 s) in Experiment 1 was longer than the longest pause duration obtained with L2-learners who were native Japanese speakers (Yamashita et al., 2014). The speech stimuli were rated on

23 items (see **Table 2.1**), and factor analysis was performed over the ratings.

Four evaluation items (i.e., “clear-cut”, “easy to understand”, “friendly”, “natural”) were directly used from Uchida (2005), and they were originally written in Japanese. The present author, who is a native-speaker of Chinese, and speaks English and Japanese, a university teacher, who is a native-speaker of Japanese, and speaks English, together with another university teacher, who is a native-speaker of Dutch, and speaks English and Japanese, collaborated to make the other 6 evaluation items (i.e., “with appropriate pause duration”, “with appropriate rhythm”, “intelligible”, “speedy”, “rushed”, and “fast”) modifying some items used by Uchida (2005) to make them more suitable for rating speech style. They also added 13 new evaluation items (e.g., “shrill”, “polite”) that were not used in Uchida (2005). These evaluation items were originally written in English.

Each item should have been translated from the original language into the other two languages among English, Chinese, and Japanese. Translations were first made by the present author, and then the two university teachers checked them. A student, who was a native speaker of Chinese, confirmed the final Chinese translations. After that, the wording of the translations was checked by at least one new native speaker in each language. (Translations are shown in **Appendix F**)

The present author made the translation based on the following procedures:

(1) Japanese evaluation items were translated to the target languages consulting reliable dictionaries, meanwhile, the items translated into the target languages were translated back to Japanese [English: Taishukan’s Unabridged Genius English-Japanese Dictionary, (Konishi & Minamide, 2001); Shogakukan Progressive Japanese-English Dictionary (4th edition),

(Kondo, & Takano, 2002); Chinese: Japanese-Chinese Dictionary (3rd edition), (University of International Business and Economics, Shogakukan, & Beijin Commercial Press, 2015); Chinese-Japanese Dictionary (3rd edition), (Beijin Commercial Press, & Shogakukan, 2016)].

(2) The translated words or phrases should be fitted in the following sentences.

This speech is _____. (Chinese: 这条语音是_____。Japanese; この話すことは_____です。)

This way of speaking is_____. (Chinese; 这种说话方式是_____。Japanese; この話し方は_____です。)

Table 2.1. Evaluation items used in Experiment 1 and Experiment 2, as judged by native-English speakers and non-native speakers (native-Chinese and native-Japanese).

Experiment 1	**Experiment 1, Experiment 2**
"intelligible", "polite", "dynamic", "clear-cut", "elegant", "smooth", "nervous", "experienced", "shrill", "fluent", "easy to understand"	"with appropriate rhythm", "rushed", "natural", "rough-timbred", "skillful", "speedy", "at a suitable tempo", "well-practiced", "fast", "with appropriate pause duration", "friendly", "high-pitched"

2.2.1 Participants

Both non-native English participants (Chinese-native speakers, Japanese-native speakers) and native-English participants joined the experiment. The native-English group consisted of 19 participants (5 males, 18-23 years old, average 20.8, SD= 1.9 years old; 14 females, 18-45 years old, average 22.5, SD= 6.8 years old). They were students

or employees from the School of Psychology, National University of Ireland, Galway, Republic of Ireland. The Irish participants were English-educated from birth.

The group of non-native participants consisted of Chinese and Japanese speakers. Data were collected from 20 native-Chinese speakers (6 males, 19-33 years old, average 23.3, SD= 4.5 years old; 14 females, 18-27 years old, average 22.2, SD= 2.3 years old). They were undergraduate students and graduate students from 8 different universities in Beijing, People's Republic of China (i.e., Peking University, University of International Relations, University of Science and Technology Beijing, University of Chinese Academy of Sciences, University of International Business and Economics, Beijing Jiaotong University, China University of Mining and Technology in Beijing, and Beijing Forestry University). Their majors varied from psychology, linguistics, civil engineering, cellular biology, to (applied) mathematics. They had studied English as their second language (L2) from the age of 6 to 16 years. Three had scores on the Test of English as a Foreign Language (TOEFL IBT; scores= 82-112), one had a score on the International English Language Testing System (IELTS; scores= 6.5), while 17 had taken the College English Test (CET-4; scores= 452-600, CET-6; scores= 450-632). One participant had scores on two different English proficiency tests. All except one had received additional English lectures in university.

The group of native-Japanese speakers consisted of 19 participants. They were students from Kyushu University, Fukuoka, Japan (13 males, 21-30 years old, average 23.8, SD= 2.52 years old; 6 females, 21-38 years old, average 25.2, SD= 5.8 years old). Five had taken TOEIC (scores= 450-895), one had taken IELTS (score= 7.0), three had taken TOEFL (two standard tests, scores= 350 and 450; one TOEFL ITP, score= 520). One Japanese participant had scores on two different English proficiency tests. Eleven of

them had not taken any English proficiency test, but had passed the entrance exam of Kyushu University, Fukuoka, Japan, which included an English proficiency test.

All participants reported to have normal hearing. Before starting the experiment, the procedure of the experiment was explained to them. All agreed to participate and had provided written informed consent. The participants were paid for their time. The experiment was conducted with prior approval of the Ethics Committee of Kyushu University, Fukuoka, Japan; the Research Ethics Committee of the National University of Ireland, Galway; and the Human Subject Review Committee of Peking University.

2.2.2 Speech segments

The segments used in Experiment 1 and Experiment 2 are shown in Table 1. 1. Four English speech segments uttered by native-English speakers were extracted from English textbooks and utilized as speech materials. We chose written materials in order to be able to systematically control the stimulus conditions.

Speech Segment 1 (Faculty of Liberal Arts, University of Tokyo English Subcommittee, 1998) was uttered by a male speaker, reading Patrick McGrath's "O'Malley and Schwartz": "*His hair hangs about his hollow, stubbled cheeks in a mess of tangled knots, and as he peers about him into the jostling throng there is in his deep-set eyes an expression of such melancholy, such sheer pain, that you would think some ghastly tragedy had befallen him, to bring him to these dire straits.*" (Patrick McGrath: "O'Malley and Schwartz").

Speech segment 2 (Faculty of Liberal Arts, University of Tokyo English Subcommittee, 1998) was also extracted from this English text book. The segment was

also uttered by a male speaker reading Gregory Bateson's “What Science Can and Cannot Predict”: “*According to the popular image of science, everything is, in principle, predictable and controllable; and if some event or process is not predictable and controllable in the present state of our knowledge, a little more knowledge and, especially, a little more know-how will enable us to predict and control the wild variables.*”

Speech Segment 3 was extracted from another English textbook with a compact disc (Faculty of Liberal Arts, University of Tokyo English Subcommittee, 2000). It was uttered by a female voice reading Mary Catherine Bateson's “Against Focused Attention”: “*Life is complicated. It is simplifying but dangerous to have one overriding concern that makes others unimportant --- rage or passion or the kind of religious exultation that seeks or inflicts martyrdom. The most striking cause of narrowed attention at the national level is warfare. In a complex world of conflicting priorities, going to war can be a tremendous relief.*”

Speech Segment 4 (Faculty of Liberal Arts, University of Tokyo English Subcommittee, 1998) was extracted from an English textbook for university students, which was accompanied by a compact disc with spoken texts. The segment was uttered by a male speaker reading Gregory Bateson's “What Science Can and Cannot Predict”: “*Under tension, a chain will break at its weakest link. That much is predictable. What is difficult is to identify the weakest link before it breaks. The generic we can know, but the specific eludes us. Some chains are designed to break at a certain tension and at a certain link. But a good chain is homogeneous, and no prediction is possible.*”

The total durations of the four speech segments were 21.02 s, 23.02 s, 31.72 s, and 29.92 s, respectively. Speech Segments 1, 2, and 3 were used in Experiment 1 (Chapter 2) and Segments 3 and 4 were used in Experiment 2 (Chapter 3). Table 1 shows the

comma- and period-pause durations for each segment. A comma pause is the pause at punctuation marks within sentences, like a comma, a semicolon, or a dash. A period pause is the pause at punctuation marks between sentences, like a period, or a question mark. The “Others” category in Table 1 are pauses mainly made for breathing. The mean pause duration for commas ranged from 0.51 s to 0.78 s, while the mean pause duration for periods ranged from 1.40 s to 1.43 s. The number of syllables ranged from 53 to 62, and the articulation rate of original speech segments varied from 3.04-3.96 syllables per second. The pause durations were comparable to the durations of commas and periods mentioned in previous studies (O’Connell & Kowal, 1986; Yamashita & Fuyuno, 2015; Liu et al., 2016; Yamashita et al., 2019). The articulation rate of the speech segments used here was a little slower than that for (American) English in daily conversation [4.88 syllables/s, (Kuhnert and Antolík, 2018); 5.12 syllables/s, (Jacewicz et al., 2009)].

Table 2.2 Speech segments used in Experiment 1 and Experiment 2.

Segment 1 (Exp 1)	**Author and Title:** Patrick McGrath's "O'Malley and Schwartz"				
Content: "His hair hangs about his hollow, stubbled cheeks in a mess of tangled knots, and as he peers about him into the jostling throng there is in his deep-set eyes an expression of such melancholy, such sheer pain, that you would think some ghastly tragedy had befallen him, to bring him to these dire straits."					
Speaker	**Number of Words**	**Number of Syllables**	**Number of Consonants**	**Average Pause Duration (s) (SD)**	**Segment Duration (s)**
male	56	72	129	Commas: 0.55 (0.23)	21.02
Segment 2 (Exp 1)	**Author and Title:** Gregory Bateson's "What Science Can and Cannot Predict"				
Content: "According to the popular image of science, everything is, in principle, predictable and controllable; and if some event or process is not predictable and controllable in the present state of our knowledge, a little more knowledge and, especially, a little more know-how will enable us to predict and control the wild variables."					
Speaker	**Number of Words**	**Number of Syllables**	**Number of Consonants**	**Average Pause Duration (s) (SD)**	**Segment Duration (s)**
male	53	91	140	Commas: 0.78 (0.33)	23.02
Segment 3 (Exps 1, 2)	**Author and Title:** Mary Catherine Bateson's "Against Focused Attention"				
Content: "Life is complicated. It is simplifying but dangerous to have one overriding concern that makes others unimportant --- rage or passion or the kind of religious exultation that seeks or inflicts martyrdom. The most striking cause of narrowed attention at the national level is warfare. In a complex world of conflicting priorities, going to war can be a tremendous relief."					
Speaker	**Number of Words**	**Number of Syllables**	**Number of Consonants**	**Average Pause Duration (s) (SD)**	**Segment Duration (s)**
female	59	102	167	Commas: 0.60 (0.04) Periods: 1.40 (0.51) Pause Duration: 1.07 (0.44)	31.72
Segment 4 (Exp 2)	**Author and Title:** Gregory Bateson's "What Science Can and Cannot Predict"				
Content: "Under tension, a chain will break at its weakest link. That much is predictable. What is difficult is to identify the weakest link before it breaks. The generic we can know, but the specific eludes us. Some chains are designed to break at a certain tension and at a certain link. But a good chain is homogeneous, and no prediction is possible."					
Speaker	**Number of Words**	**Number of Syllables**	**Number of Consonants**	**Average Pause Duration (s) (SD)**	**Segment Duration (s)**
male	62	91	148	Commas: 0.51 (0.04) Periods: 1.43 (0.60) Pause Duration: 1.08 (0.65)	29.94

2.2.3 Speech Stimuli

Speech segments were extracted from English textbooks with a compact disc (Faculty of Liberal Arts, University of Tokyo English Subcommittee, 1998; Faculty of Liberal Arts, University of Tokyo English Subcommittee, 2000). The English textbooks were used in the University of Tokyo, Japan, for English education. The editors were native-English speakers from the Faculty of Liberal Arts, University of Tokyo English Subcommittee, and English-education professionals.

The segments were prepared as follows. First, the speech segments were transformed from the ".cda" format and saved as ".wav" files, in order to edit the waveforms. Next, sections with sound energy (i.e., utterances) and sections without sound energy (i.e., silent sections) were semi-automatically extracted using the audio-software "Praat" (Boersma & Weenink, 2015). Using "Praat", the speech segments were annotated to a TextGrid (Annotate function: to TextGrid (silences); guidelines for settings: Silence threshold: -35 dB; Minimum silent interval duration: 0.1 s; Minimum sounding: 0.1 s). All the utterances were then saved as separate digital samples. Following this, at temporal positions in the three original speech segments at which a comma, a period, a semicolon or a dash appeared, a new pause duration was inserted using 'J' programming language. Every other pause duration longer than 0.1 s was adjusted to 0 s, because we only focused on durations at punctuation marks. The pause durations that were inserted for commas and periods were fixed at 0.075, 0.15, 0.3, 0.6, 1.2, 2.4, and 4.8 s, resulting in 21 speech stimuli in total. The duration of each pause was the same for commas and for periods. The pause at the semi-colon in Speech Segment 2 spoken by the male speaker and at the dash in Speech Segment 3 spoken by the female speaker were also made with the seven

durations. Finally, the average intensities of the stimuli were equalized (65 dBA). Speech only from the left channel was used to make a mono speech sample, enabling easier calibration of the sound level before presentation to the participants. Figure 2.1 shows examples of waveforms of the stimuli used throughout this thesis. Figure 2.1 (a) shows an example of the waveforms of an original speech segment. Panels (b), (c), and (d) in Figure 2.1 show an example of the waveforms of a segment (Segment 3), in which all pause durations were manipulated to 0.15 s, 0.60 s, or 2.40 s, respectively.

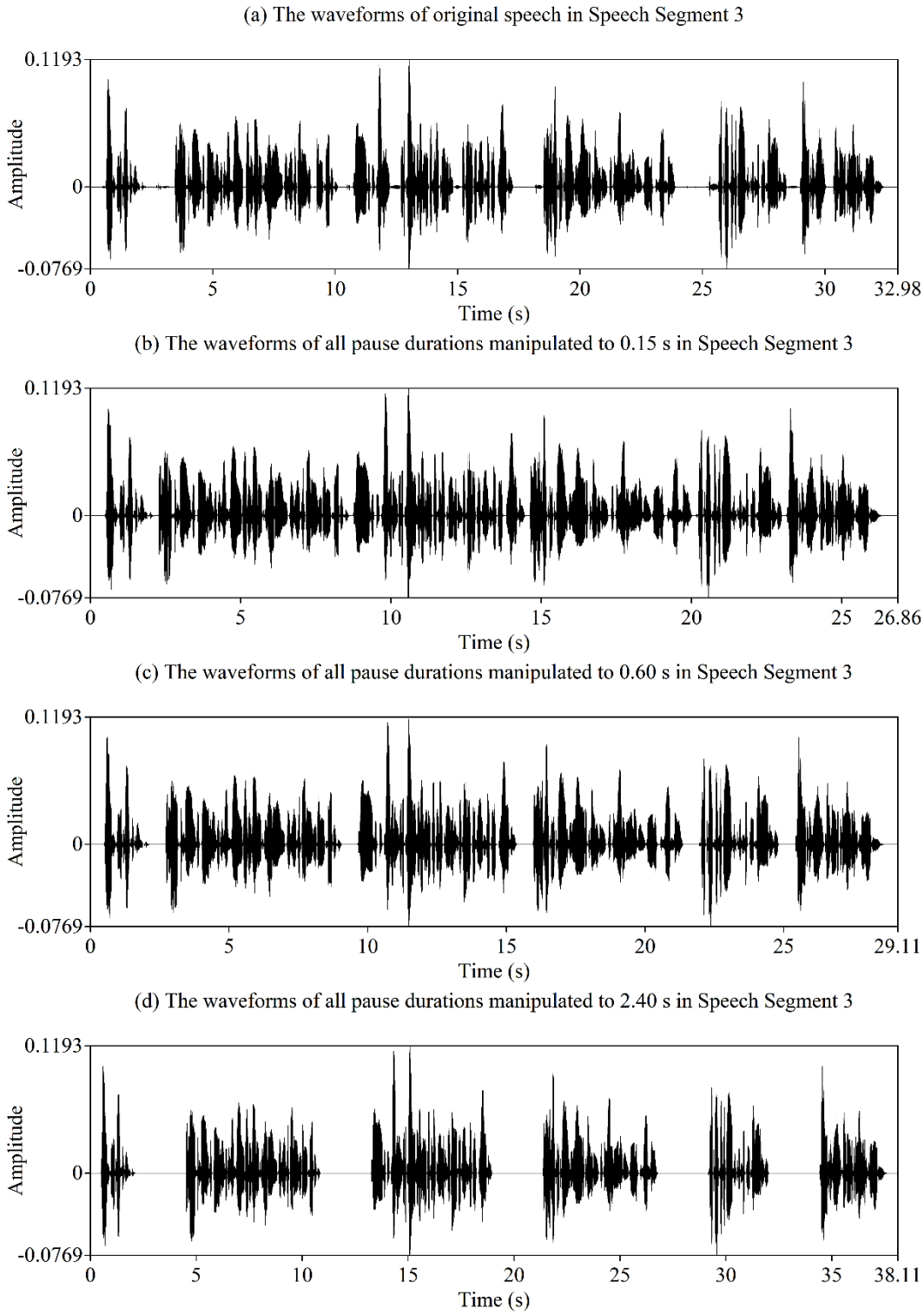


Figure 2.1 Examples of waveforms of the stimuli used throughout this thesis. (a) The waveforms of original speech of Speech Segment 3. (b), (c), and (d) The waveforms of all pause durations manipulated to 0.15 s, 0.60 s, or 2.40 s, respectively, in Speech Segment 3. (Taken from "Praat" software window.)

2.2.4 Apparatus

The speech stimuli were diotically presented to the participants in a soundproof booth (background level<30 dBA), by means of monitor headphones (Roland RH-300) and a USB headphone amplifier (AT-HA40USB). The stimuli were presented and controlled through an interface using a tablet (Microsoft Surface 3 64GB, OS Windows 8.1). A customized program in 'J'-language was used to equalize the level of the stimuli. The sound pressure level was measured with a sound level meter (ACO, Type 6240), and an artificial ear (Brüel and Kjær, 4153, Nærum, Denmark). Figure 2.2 shows the apparatus used in Experiment 1 and how they were connected to each other.

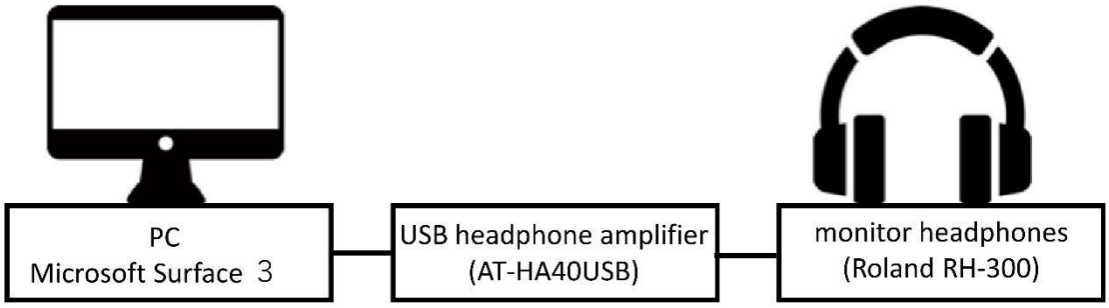


Figure 2.2 The Apparatus used in Experiment 1.

2.2.5 Procedure

The experiment was conducted in three different places. The data from the native-English participants were obtained in Galway, Republic of Ireland, the data from the Chinese participants were gathered in Beijing, People's Republic of China, while the data from the Japanese participants were obtained in Fukuoka, Japan. In the soundproof booth, the stimuli were diotically presented to the participants in three sessions. In all three sessions, the participants were asked to judge the stimuli on 23 evaluation items using a 10-point rating scale from "not" (1) to "very much" (10). The evaluation items are

indicated in **Table 2.1**. They were selected based on research on the relation between temporal structures of speech and listeners' impressions of the speaker's personality (Uchida, 2005). Six items used by Uchida (2005) were modified to make them more suitable for rating speech style by the present author and the other two university teachers. Thirteen evaluation items were newly conceived. The evaluation items, originally in Japanese and translated into English and Chinese for the speakers of those languages, consisted of 16 positive adjective words, like "fluent", "natural", and "skillful", 4 negative adjective words ("shrill", "nervous", "rushed", and "rough-timbred"), and 3 neutral/negative words ("speedy", "high-pitched", and "fast").

The stimuli were presented to the participants through headphones, 0.5 s after the participant pressed the "PLAY" button on the interface. When stimulus presentation was finished, the participants rated the stimulus on the 23 evaluation items, using pen and paper on which the 10-point rating scales were indicated. There was no time limit for participants to give each rating; the experiment was self-paced. Figure 2.3 shows the experiment interface used by the participant.

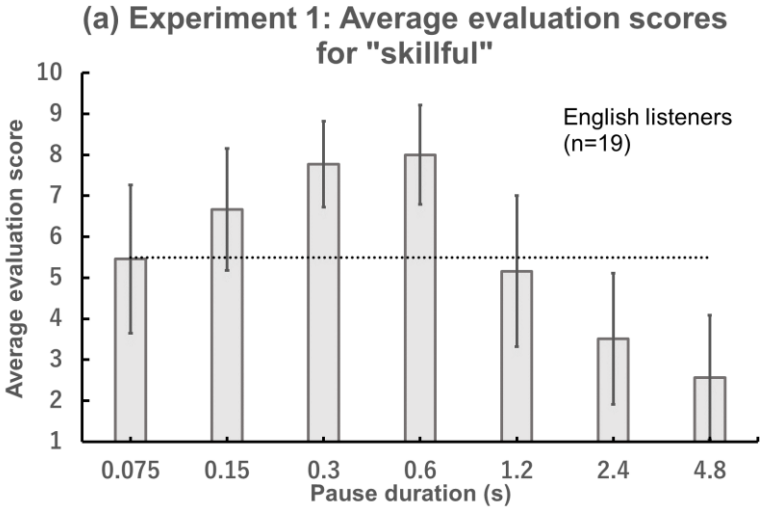
Figure 2.3 The experiment interface used in Experiment 1.

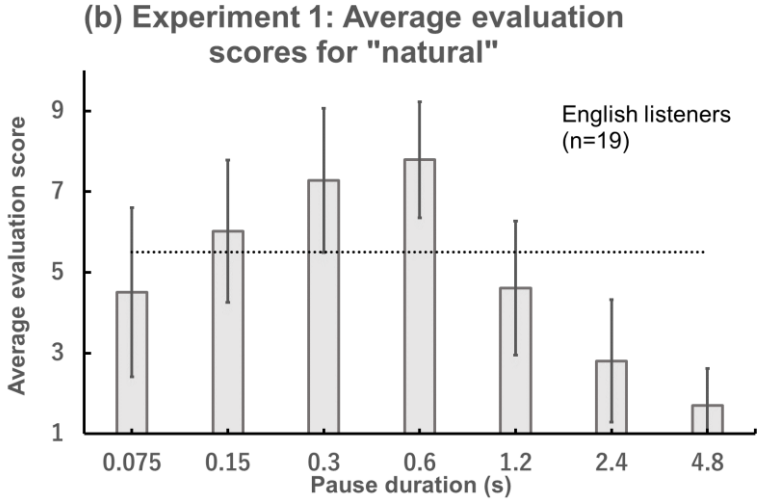
Before the experiment, there were 7 practice trials, randomly chosen from the three speech segments. The results of these practice trials were not considered for further analysis. The participants could take a break following practice. The experiment was divided into two sessions, with the second session following the first, with a break in between. There were 12 trials in the first session, and 11 trials in the second session. The first trial and the last trial in each session were the same, but the results of the first trial were not analyzed. In total, rating data were obtained from 21 speech stimuli (3 speech segments ×7 durations). The experiment took approximately 50 minutes. After the last session, the participants were asked to fill in a questionnaire about their personal details and language background.

2.3 Results

2.3.1 Average evaluation scores for each item

First of all, to test whether some evaluation items could be grouped to show the same tendencies related to the change of pause duration, the distributions of the original data obtained from Experiment 1 were directly observed. The tendencies of change in the data could be classified roughly into three types of distributions according to the average evaluation scores of each evaluation item, i.e., a "curve type" (see **Figure 2.4**), which had relatively low evaluation scores with short or long pause durations, but relatively high evaluation scores for intermediate pause durations; an approximately "steady decrease type" (see **Figure 2.5**), in which the evaluation scores decreased steadily as the pause duration became longer; and a "no obvious change type", in which the evaluation scores did not change obviously as the pause duration changed (see **Figure 2.6**). Interestingly, this could be observed also in the Chinese language group (see **Figures 2.7**), and in the Japanese language group (see **Figures 2.8**). From this, it could be expected that according to the type of evaluation items, pause duration affected the perception of the speech differently. (Average evaluation scores for each evaluation item are shown in **Appendix E**). Thus, we decided to use factor analysis to see whether we could summarize these tendencies into factors.


(a) Experiment 1: Average evaluation scores for "skillful"
Average evaluation score
10
9
8
7
6
5
4
3
2
1
English listeners
(n=19)
0.075
0.15
0.3
0.6
1.2
2.4
4.8
Pause duration (s)


(b) Experiment 1: Average evaluation scores for "natural"
Average evaluation score
9
7
5
3
1
English listeners
(n=19)
0.075
0.15
0.3
0.6
1.2
2.4
4.8
Pause duration (s)

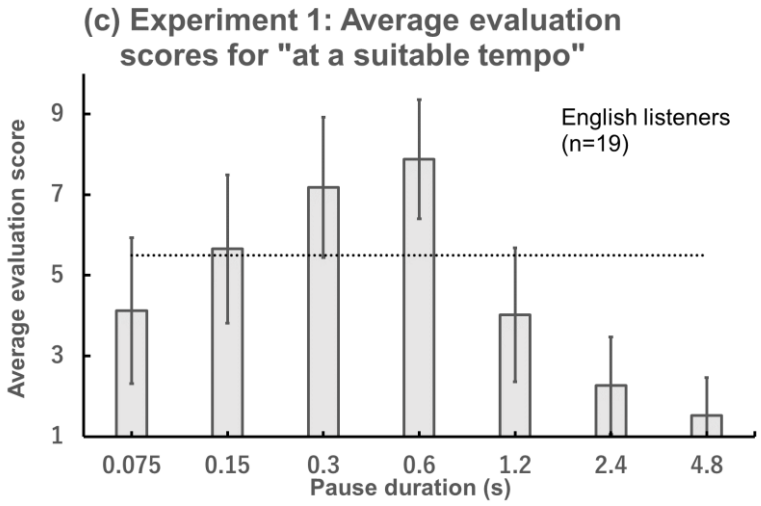
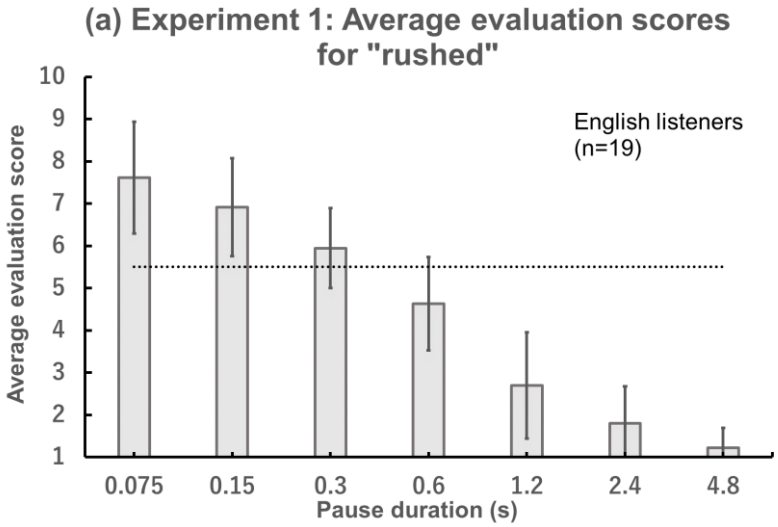
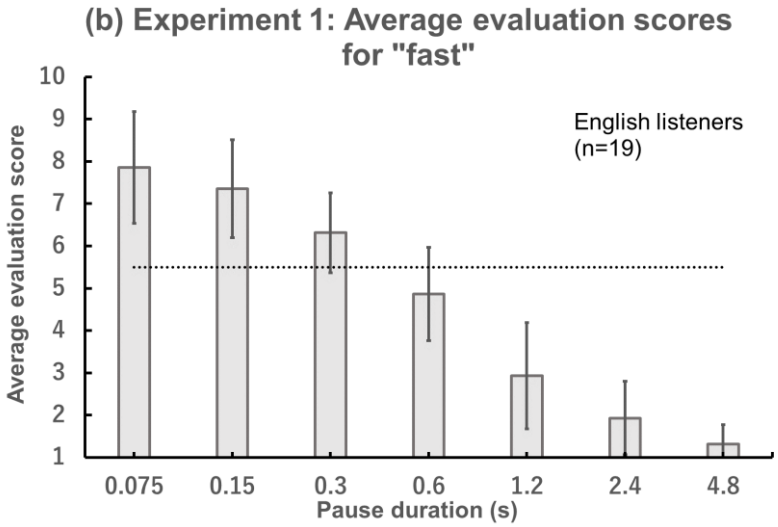


Figure 2.4 Examples of a “curve type” distribution in the evaluation scores of the English group. (a) Average evaluation scores for the item “skillful”. (b) Average evaluation scores for “natural”. (c) Average evaluation scores for “at a suitable tempo”. The error bar shows the standard deviation. The black dotted line shows the scale midpoint.


(a) Experiment 1: Average evaluation scores for "rushed"
English listeners (n=19)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
0.075
0.15
0.3
0.6
1.2
2.4
4.8
Pause duration (s)


(b) Experiment 1: Average evaluation scores for "fast"
English listeners (n=19)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
0.075
0.15
0.3
0.6
1.2
2.4
4.8
Pause duration (s)

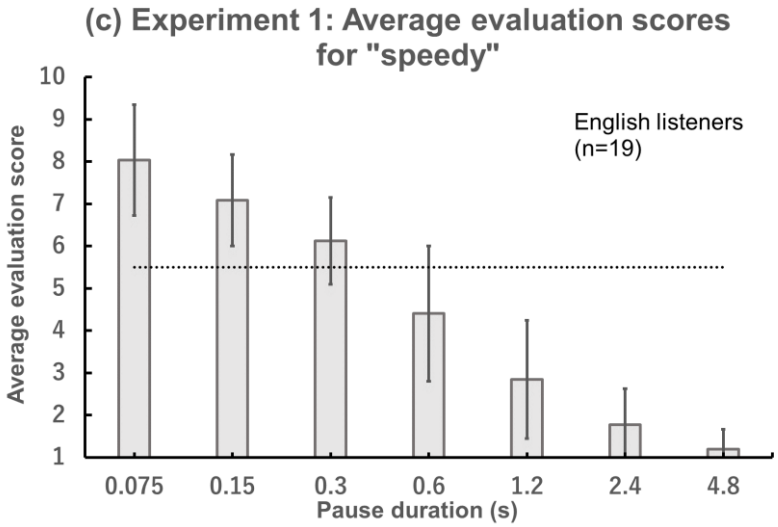
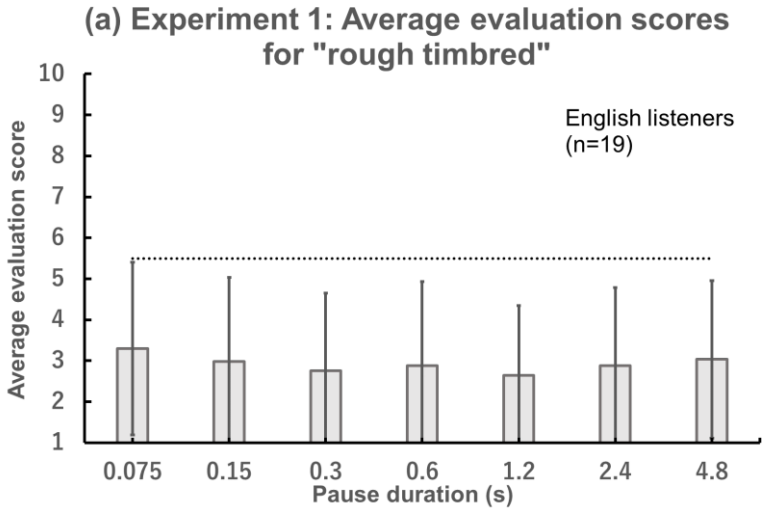
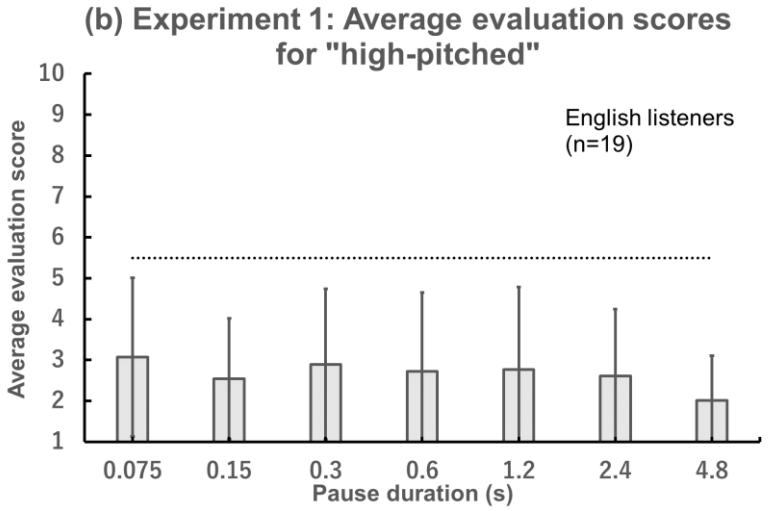


Figure 2.5 Examples of a “steady decrease type” distribution in the English group. (a) Average evaluation scores for the item “rushed”. (b) Average evaluation scores for “fast”. (c) Average evaluation scores for “speedy”. The error bars show the standard deviation. The black dotted line shows the scale midpoint.


(a) Experiment 1: Average evaluation scores for "rough timbred"
English listeners (n=19)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
0.075
0.15
0.3
0.6
1.2
2.4
4.8
Pause duration (s)


(b) Experiment 1: Average evaluation scores for "high-pitched"
English listeners (n=19)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
0.075
0.15
0.3
0.6
1.2
2.4
4.8
Pause duration (s)

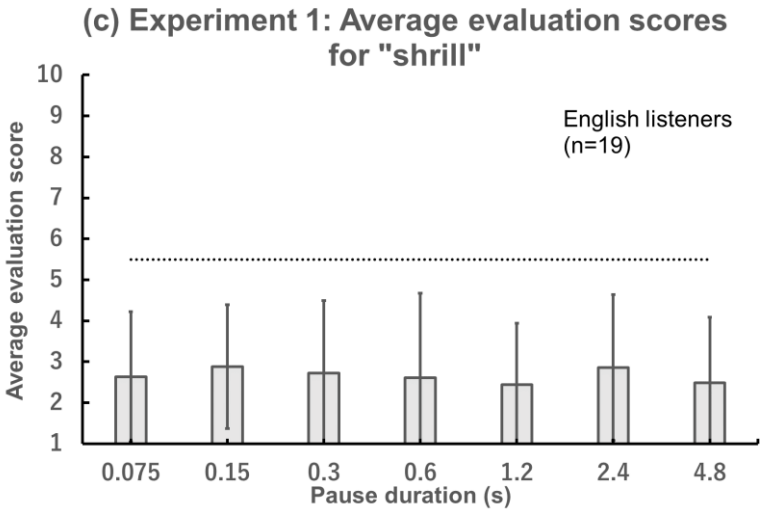
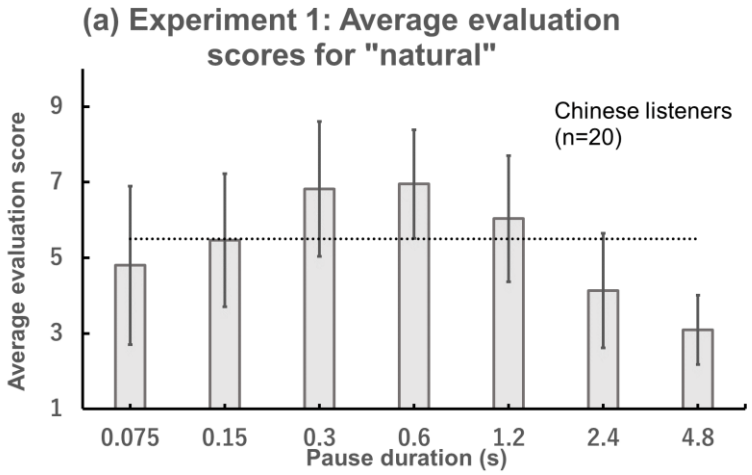
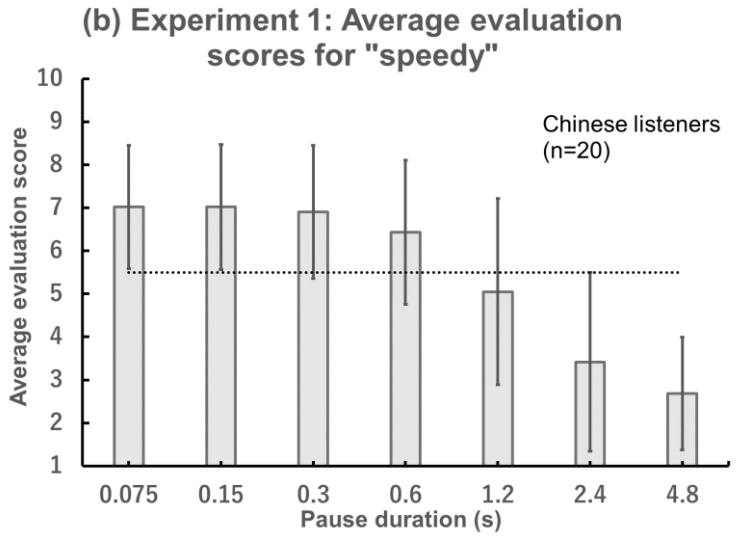


Figure 2.6 Examples of a "no obvious type" distribution in the English group. (a) Average evaluation scores for the item "rough timbred". (b) Average evaluation scores for "high-pitched". (c) Average evaluation scores for "shrill". The error bars show the standard deviation. The black dotted line shows the scale midpoint.


(a) Experiment 1: Average evaluation scores for "natural"
Chinese listeners (n=20)
Average evaluation score
9
7
5
3
1
0.075
0.15
0.3
0.6
1.2
2.4
4.8
Pause duration (s)


(b) Experiment 1: Average evaluation scores for "speedy"
Chinese listeners (n=20)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
0.075
0.15
0.3
0.6
1.2
2.4
4.8
Pause duration (s)

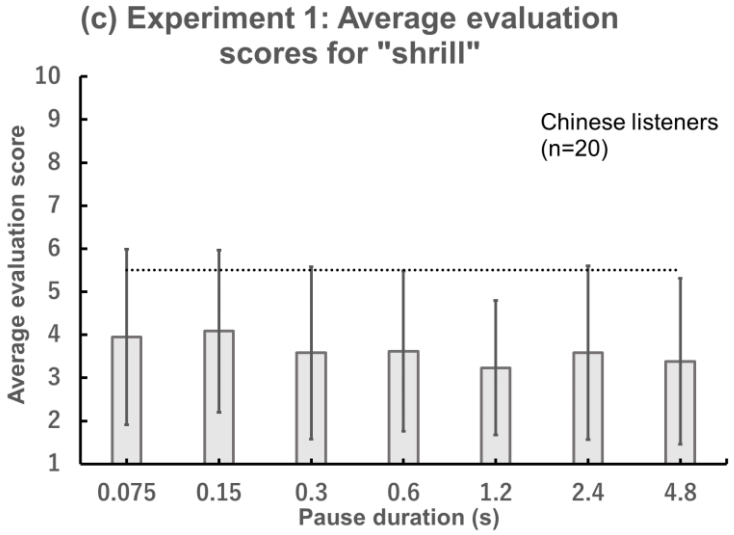
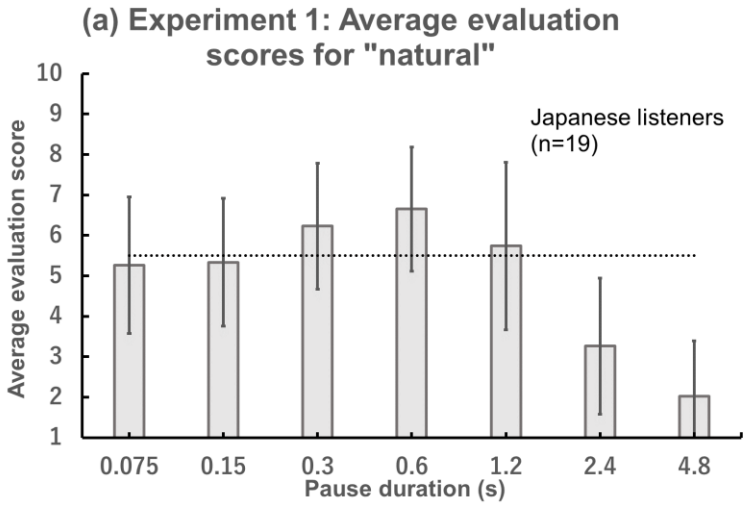
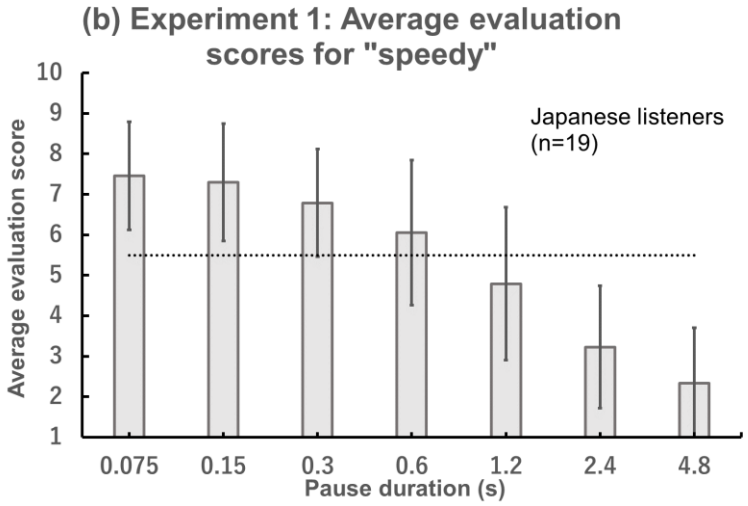


Figure 2.7 Examples of three types of distributions in the Chinese group. (a) "Curve type" distribution: Average evaluation scores for the item "natural". (b) "Steady decrease type" distribution: Average evaluation scores for "speedy". (c) "No obvious type" distribution: Average evaluation scores for "shrill". The error bars show the standard deviation. The black dotted line shows the scale midpoint.


(a) Experiment 1: Average evaluation scores for "natural"
Japanese listeners (n=19)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
0.075
0.15
0.3
0.6
1.2
2.4
4.8
Pause duration (s)


(b) Experiment 1: Average evaluation scores for "speedy"
Japanese listeners (n=19)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
0.075
0.15
0.3
0.6
1.2
2.4
4.8
Pause duration (s)

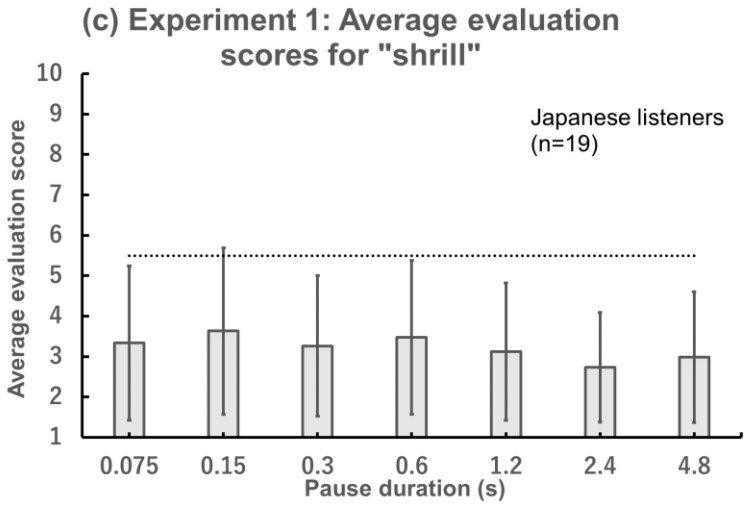


Figure 2.8 Examples of three types of distributions in the Japanese group. (a) “Curve type” distribution: Average evaluation scores for the item “natural”. (b) “Steady decrease type” distribution: Average evaluation scores for “speedy”. (c) “No obvious type” distribution: Average evaluation scores for “shrill”. The error bars show the standard deviation. The black dotted line shows the scale midpoint.

2.3.2 Results of factor analysis

The results were analyzed in the following steps. In order to check whether the rating data were suitable for factor analysis, Kaiser-Meyer-Olkin (KMO) tests were performed. The results showed that the sampling was adequate overall for the data obtained from the native-English listeners (0.947), the Chinese listeners (0.944), and the Japanese listeners (0.934). [Bartlett's tests of sphericity were all significant ($p<0.001$)]. Following principal component analysis (PCA) with varimax rotation, four factors were extracted for all three language groups. The factors were labeled according to the categorical items, following Pett et al. (2003). The first factor was called the "Speech Naturalness factor". In this factor, the evaluation items "elegant", "skillful", "smooth", "with appropriate rhythm", "natural", "experienced", "well-practiced", "with appropriate pause duration", "at a suitable tempo", "polite", "friendly", "fluent", "intelligible", and "easy to understand" were included for all three language groups. The second factor could be summarized as the "Speech Rate factor"; it included evaluation items "speedy", "rushed", and "fast" for all three language groups. The third factor ("high-pitched", "shrill") and the fourth factor ("rough-timbred") related to sound quality. The cumulative percentages of variance at the third and fourth factor were 66% and 71%, respectively, in all of the three language groups. The first (Speech Naturalness) and the second factor (Speech Rate) were taken into further consideration, because their cumulative percentage of variance was about 60% for all three language groups. Table 2.3 shows the details of the results of the factor analysis, including the total variance for the three language groups.

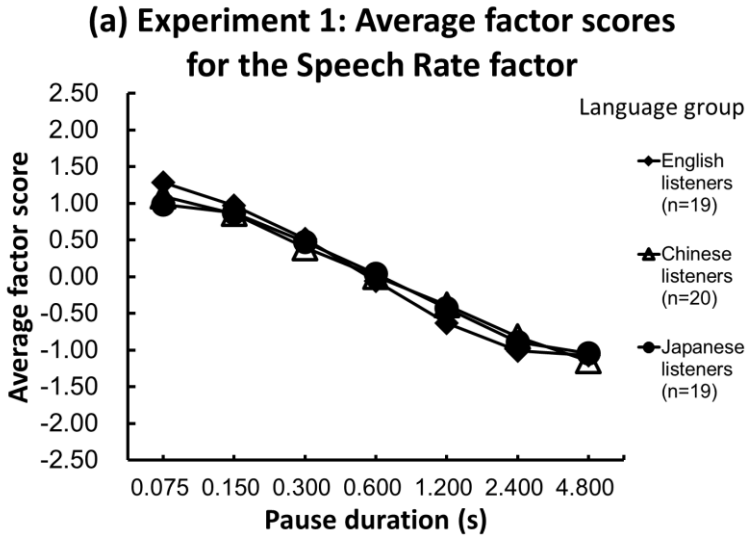
Table 2.3 Results of Experiment 1: factor analysis details.

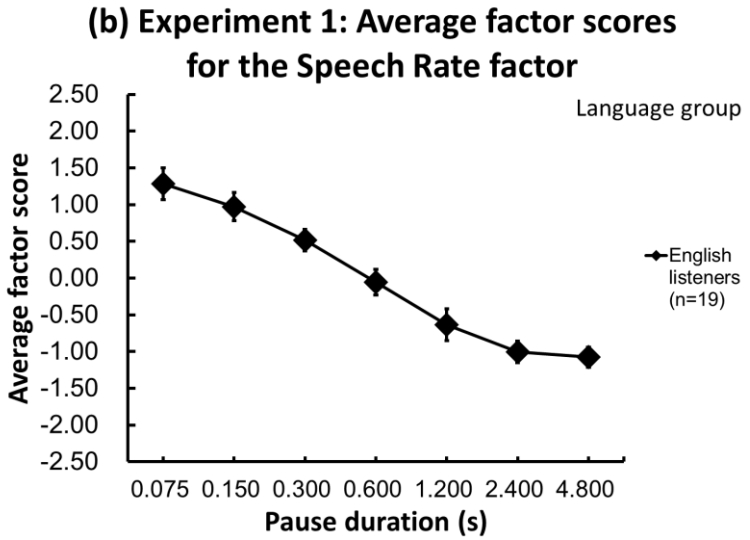
Language Group	Total Variance Explained						
	Component	Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
		Total	Percentage of Variance	Cumulative Percentage	Total	Percentage of Variance	Cumulative Percentage
English listeners	1	11.259	48.953	48.953	9.925	43.153	43.153
	2	2.569	11.171	60.124	3.829	16.646	59.799
	3	1.665	7.239	67.363	1.656	7.200	67.000
	4	1.025	4.457	71.820	1.109	4.821	71.820
Chinese listeners	1	11.186	48.633	48.633	10.036	43.633	43.633
	2	3.309	14.386	63.019	3.411	14.832	58.465
	3	1.365	5.936	68.954	1.925	8.371	66.836
	4	1.087	4.728	73.682	1.575	6.846	73.682
Japanese listeners	1	10.410	45.260	45.260	8.703	37.838	37.838
	2	3.262	14.184	59.444	4.756	20.677	58.515
	3	1.609	6.996	66.440	1.800	7.825	66.340
	4	1.071	4.658	71.097	1.094	4.757	71.097

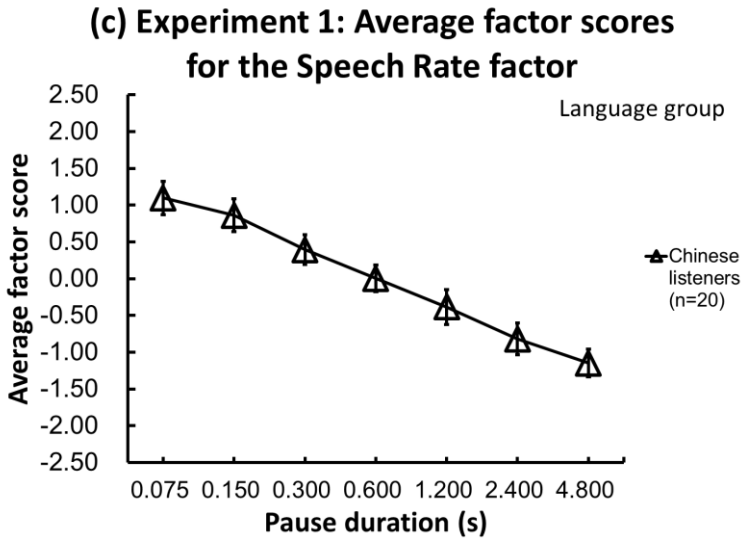
2.3.3 Results of Speech Rate factor

Figure 2.9 shows the average factor scores for the Speech Rate factor. Since Shapiro-Wilk tests showed that the factor scores were not normally distributed for all three language groups, comparisons of factor scores were performed with Friedman tests ($p<0.05$), followed by pair-wise Wilcoxon tests with Holm-Bonferroni correction for multiple comparisons. For all three language groups, the Friedman tests were significant [native-English group ($\chi 2$ (df= 6, n= 19) = 104.4, $p<0.0001$; Chinese group ($\chi 2$ (df= 6, n= 20) = 111.3, $p<0.0001$; Japanese group ($\chi 2$ (df= 6, n= 19) = 97.3, $p<0.0001$]. Overall, paired comparisons showed that the factor scores significantly decreased as pause duration increased. There were only two exceptions. The difference between the factor

scores for the stimuli with the 2.4-s and the 4.8-s pause durations was not significant in the native-English group, while in the Japanese language group, the difference between the stimuli with the 0.075-s and the 0.15-s pause durations was not significant. The Kendall's Coefficient of Concordance test showed that the factor scores obtained for the three language groups were highly similar (Kendall's $W= 1.00$, $p<0.01$, $n= 3$, $k= 7$).


(a) Experiment 1: Average factor scores for the Speech Rate factor
Average factor score
2.50
2.00
1.50
1.00
0.50
0.00
-0.50
-1.00
-1.50
-2.00
-2.50
0.075 0.150 0.300 0.600 1.200 2.400 4.800
Pause duration (s)
Language group
English listeners (n=19)
Chinese listeners (n=20)
Japanese listeners (n=19)


(b) Experiment 1: Average factor scores for the Speech Rate factor
Average factor score
2.50
2.00
1.50
1.00
0.50
0.00
-0.50
-1.00
-1.50
-2.00
-2.50
0.075 0.150 0.300 0.600 1.200 2.400 4.800
Pause duration (s)
Language group
English listeners (n=19)



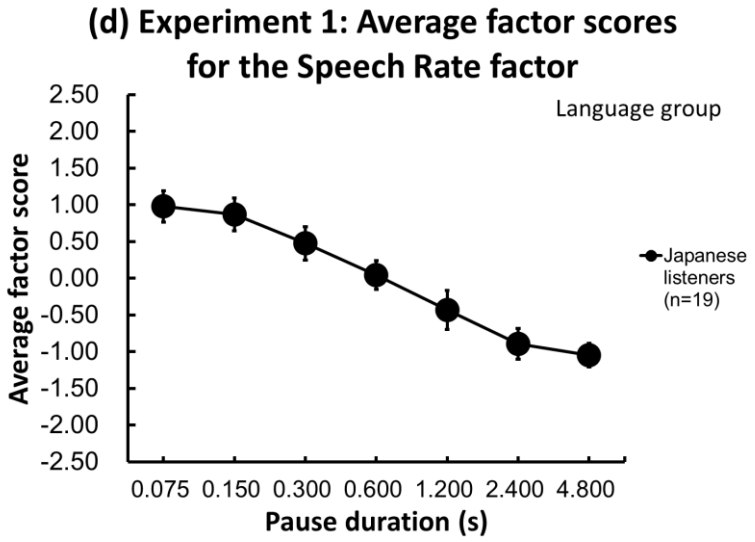
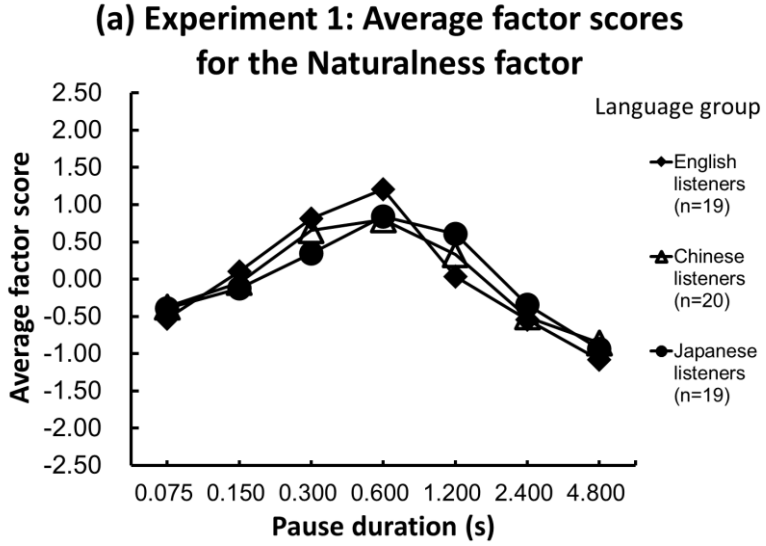
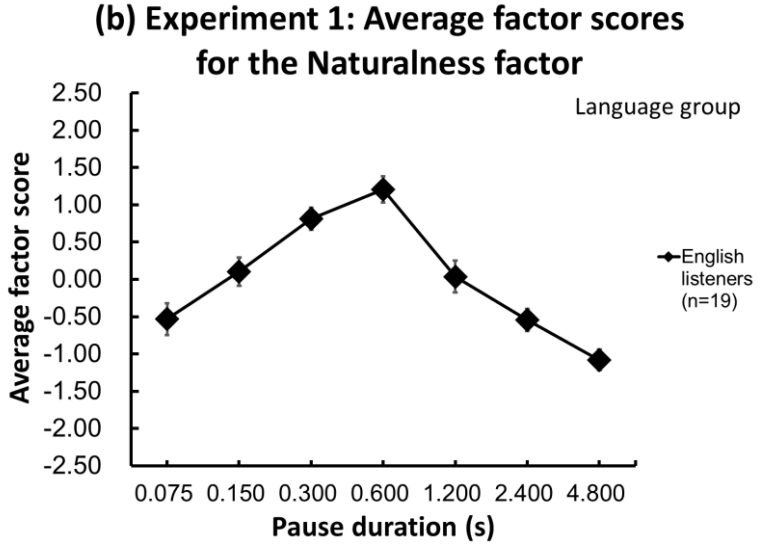


Figure 2.9 Results of Experiment 1. (a) The average factor scores for the Speech Rate factor from the three language groups. (b) The average factor scores for the Speech Rate factor from the English listeners. (c) The average factor scores for the Speech Rate factor from the Chinese listeners. (d) The average factor scores for the Speech Rate factor from the Japanese listeners. The error bars show the 95%-confidence intervals.

2.3.4 Results of Naturalness factor

Figure 2.10 shows the average factor scores for the Naturalness factor. The Kendall's Coefficient of Concordance test showed that the factor scores for the three language groups were very similar for this factor as well (Kendall's $W = 0.94$, $p<0.01$, $n = 3$, $k = 7$). Since the factor scores for the native-English group were not normally distributed, again Friedman tests with Holm-Bonferroni correction were performed over factor scores. For the Naturalness factor the test results were significant for all three language groups [native-English group ($\chi 2$ ($df = 6$, $n = 19$) = 93.3, $p<0.0001$; Chinese group ($\chi 2$ ($df = 6$, $n = 20$) = 92.5, $p<0.0001$; Japanese group ($\chi 2$ ($df = 6$, $n = 19$) = 73.9, $p<0.0001$]. For the native-English group, this factor score was significantly higher than that for any of the other stimuli. For the Chinese group, only the factor score for the 0.3-s stimuli was not significantly lower than that for the 0.6-s stimuli. For the Japanese group, the factor score for the 0.6-s stimuli was not significantly higher than that for the 1.2-s stimuli. In conclusion, the Naturalness factor scores for the stimuli with the 0.6-s pause duration were the highest in all three groups.


(a) Experiment 1: Average factor scores for the Naturalness factor
Average factor score
2.50
2.00
1.50
1.00
0.50
0.00
-0.50
-1.00
-1.50
-2.00
-2.50
0.075 0.150 0.300 0.600 1.200 2.400 4.800
Pause duration (s)
Language group
English listeners (n=19)
Chinese listeners (n=20)
Japanese listeners (n=19)


(b) Experiment 1: Average factor scores for the Naturalness factor
Average factor score
2.50
2.00
1.50
1.00
0.50
0.00
-0.50
-1.00
-1.50
-2.00
-2.50
0.075 0.150 0.300 0.600 1.200 2.400 4.800
Pause duration (s)
Language group
English listeners (n=19)

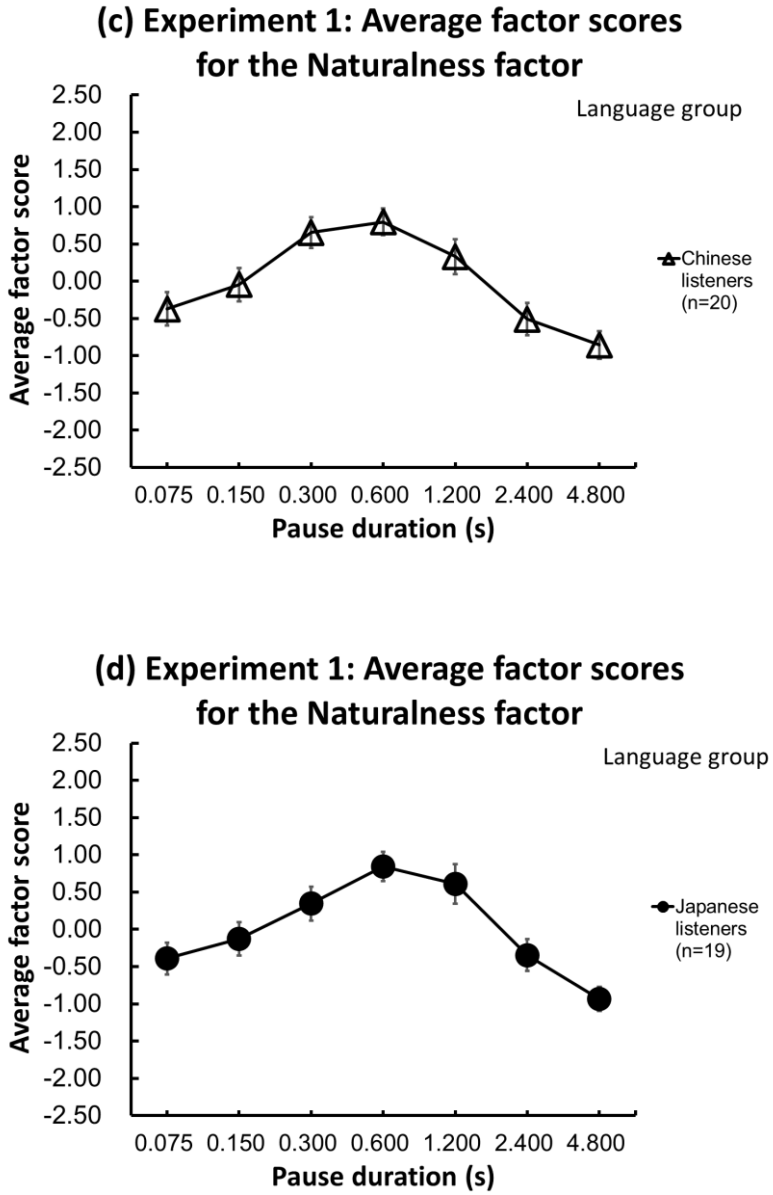


Figure 2.10 Results of Experiment 1. (a) The average factor scores for the Naturalness factor from the three language groups. (b) The average factor scores for the Naturalness factor from the English listeners. (c) The average factor scores for the Naturalness factor from the Chinese listeners. (d) The average factor scores for the Naturalness factor from the Japanese listeners. The error bars show the 95%-confidence intervals.

2.4 Discussion

The same two factors were shown in all three language groups. According to the contents of evaluation items (i.e., “speedy”, “rushed”, and “fast” in one factor; “elegant,” “skillful,” “smooth,” “with appropriate rhythm,” “natural,” “experienced,” “well-practiced,” “with appropriate pause duration,” “at a suitable tempo,” “polite,” “friendly,” “fluent,” “intelligible,” and “easy to understand” in another factor) that were grouped in all three language groups, the participants must have perceived speech rate and speech naturalness.

Factor analysis over the rating data revealed two noteworthy tendencies. First, for all three language groups, the factor scores for the Speech Rate factor (see **Figure 2.9**) decreased as pause duration increased. Although the physical speech rate (i.e., the articulation rate) of the utterances used in Experiment 1 was the same, the listeners perceived a decrease in the overall speech rate with an increase only in pause duration. The results were in line with Uchida (2005): The evaluation scores were higher if the overall physical speech rate was higher. In the preliminary research with Japanese L2-learners of English described in the introduction (Liu et al., 2016), a significant negative correlation between speech rate and pause duration was found. The present results must be closely connected to this.

As for the Naturalness factor (see **Figure 2.10**), the pause duration of 0.6 s received the highest factor scores. However, there were slight differences between language groups. For example, the difference between the 0.6-s stimuli and the other pause duration conditions was more pronounced for the native-English group than for the non-native groups. Factor scores for stimuli with relatively short (0.075 s and 0.15 s) and long (2.4 s and 4.8 s) pause durations received the lowest scores. In Uchida (2005), the original block

got the highest evaluation scores in Speech Naturalness, but the fast and slow block got relatively lower scores. The present results were quite similar.

The results of Experiment 1 thus suggest that just changing the pause duration for commas and periods in the same duration can change the subjective impression systematically, affecting the perceived speech rate and speech naturalness. A pause duration of 0.6 s seemed to make speech the most natural for both native and non-native English listeners. This duration is also a good index for music tempo (Fraisse, 1982), suggesting a commonality in the perception of temporal properties of music and speech. We anticipate these results would be useful in training L2-learners how long and when to pause, and let them become confident in controlling the timing and rhythm.

One limitation of the present experiment, however, was that the pause duration was fixed for each punctuation mark, while previous research has shown that the physical duration of periods in spoken texts is approximately twice as long as that of commas, as mentioned in the introduction (O'Connell & Kowal, 1986; Yamashita & Fuyuno, 2015; Liu et al., 2016; Yamashita et al., 2019). Therefore, in order to further investigate how pause duration influences the subjective impression of speech, in Experiment 2, the comma- and the period duration were varied independently.

Chapter 3 – Experiment 2

3.1 Introduction

The results of Experiment 1 showed that the factor scores of the naturalness of speech were the highest when the pause durations at punctuation marks were manipulated into 0.6 s. On the other hand, the factor scores of the rate of the speech decreased steadily as the manipulated pause duration became longer. These results could be observed in both the native English speakers and non-native (Chinese and Japanese) English speakers.

However, several studies showed that pause durations between sentences (i.e., periods) were longer than within sentences (i.e., commas; Cruttenden, 1986). For example, in an analysis of the verbal and non-verbal performances of public speaking from English speakers, the average pause duration was 0.49 s within sentences, and 1.01 s between sentences (Yamashita and Fuyuno, 2015). Interestingly, in a study of oral deliveries of sermons in German, the mean duration of commas was 0.47 s, and the mean duration of periods was 0.98 s (O'Connell and Kowal,1986). Thus, the mean duration of periods was typically twice as long as that in commas. Therefore, in the second listening experiment, the pause duration in commas and periods were varied independently.

The purpose of Experiment 2 was to investigate how subjective impression would change when the manipulated pause duration for commas and that for periods were varied independently. Both native and non-native speakers of English (Chinese and Japanese) were employed to investigate whether they still preferred a similar pause duration as the most favorable pause duration.

3.2 Method

In Experiment 2, comma- and period-pause durations were manipulated separately, in 7 steps varying from 0.15 to 2.4 s. In this experiment, the original speech and speech without pauses were also included, as control conditions. Similar to Experiment 1, the participants were native and non-native English speakers (native-Japanese and native-Chinese speakers). We investigated how the listeners' impressions would change as a function of pause duration by collecting rating scale data for 12 evaluation items (see **Table 2.1**), which were then subjected to factor analysis.

3.2.1 Participants

Three participant groups consisted of native-English speakers and Chinese and Japanese non-native English speakers. The native-English group consisted of 24 participants (18 students from the National University of Ireland, Galway, Republic of Ireland, and 6 students or English-education professionals in Fukuoka, Japan). They were 14 males (18-49 years old, average 27.0, SD= 9.6) and 10 females (19-39 years old, average 22.2, SD= 5.7). The Irish participants were English-educated from birth, and one of them had participated in Experiment 1. The Chinese non-native group consisted of 20 native-Mandarin Chinese speakers. They were students from Kyushu University, Fukuoka, Japan (8 males, 23-34 years old, average 26.5, SD= 3.1; 12 females, 19-26 years old, average 23.8, SD= 1.7). The Japanese non-native group consisted of 20 native-Japanese speakers. They were also students from Kyushu University, Fukuoka, Japan (10 males, 21-25 years old, average 22.6, SD= 1.4; 10 females, 20-22 years old, average 21.6, SD= 0.7).

Out of the 20 Chinese participants, 4 had scores on the Test of English as a

Foreign Language (TOEFL IBT; scores= 85-99), 8 had scores on the Test of English for International Communication (TOEIC; scores= 630-885), 4 had scores on the International English Language Testing System (IELTS; scores= 6.0-7.5), while 14 had taken the College English Test (CET-4; scores= 440-500, CET-6; scores= 450-600). Eight Chinese participants had scores on two different English proficiency tests, while one had three different English certificates. None of them had participated in Experiment 1. From the 20 Japanese participants, 10 had taken TOEIC (scores= 480-895), one had taken IELTS (score= 7.0), one had taken TOEFL ITP (score= 500), and 4 had completed TOEFL (scores= 400-600). Two students had taken two tests, while 6 had not taken any English proficiency test yet, but had passed the entrance exam of Kyushu University, Fukuoka, Japan, which includes an English proficiency test. Two of them had participated in Experiment 1. All participants reported to have normal hearing, and were paid for their time. All agreed to participate and provided written informed consent, after the procedure of the experiment was explained to them. The experiment was conducted with prior approval of the Ethics Committee of Kyushu University, Fukuoka, Japan and the Research Ethics Committee of the National University of Ireland, Galway.

3.2.2 Speech Stimuli and Apparatus

Two English speech segments (Speech Segment 3 and Speech Segment 4, **Table 2.2**) were selected as stimuli. One speech segment was the same as in Experiment 1 (Speech Segment 3), spoken by a female speaker. The other segment (Speech Segment 4) was newly extracted from “The Universe of English II” (1998), which was uttered by a male speaker. The stimulus preparation was the same as in Experiment 1. The pause durations were 0.15, 0.3, 0.6, 1.2, and 2.4 s, and the comma duration and the period

duration were varied independently, resulting in 25 stimuli for each segment. Furthermore, different from Experiment 1, for both speech segments stimuli without any pauses were made for a control condition, and the original speech segments with the pause durations as uttered by the male or the female speaker were used as well. The original speech segments included other pauses where there was no punctuation mark. In total, 54 stimuli were used in the experiment, and the average presentation levels of the stimuli were equalized (65 dBA). **Figure 3.1** shows examples of waveforms of the stimuli: (a) The waveforms of original speech in Speech Segment 3. (b) The waveforms of all pause durations manipulated to 0 s in Speech Segment 3. (c) Speech Segment 3 with a comma-pause duration of 0.30 s, and a period-pause duration of 0.60 s. (d) Speech Segment 3 with a comma-pause duration of 0.60 s, and a period-pause duration of 1.20 s. (Taken from “Praat” software window.) The same apparatus was used as in Experiment 1 (see **Figure 2.2**).

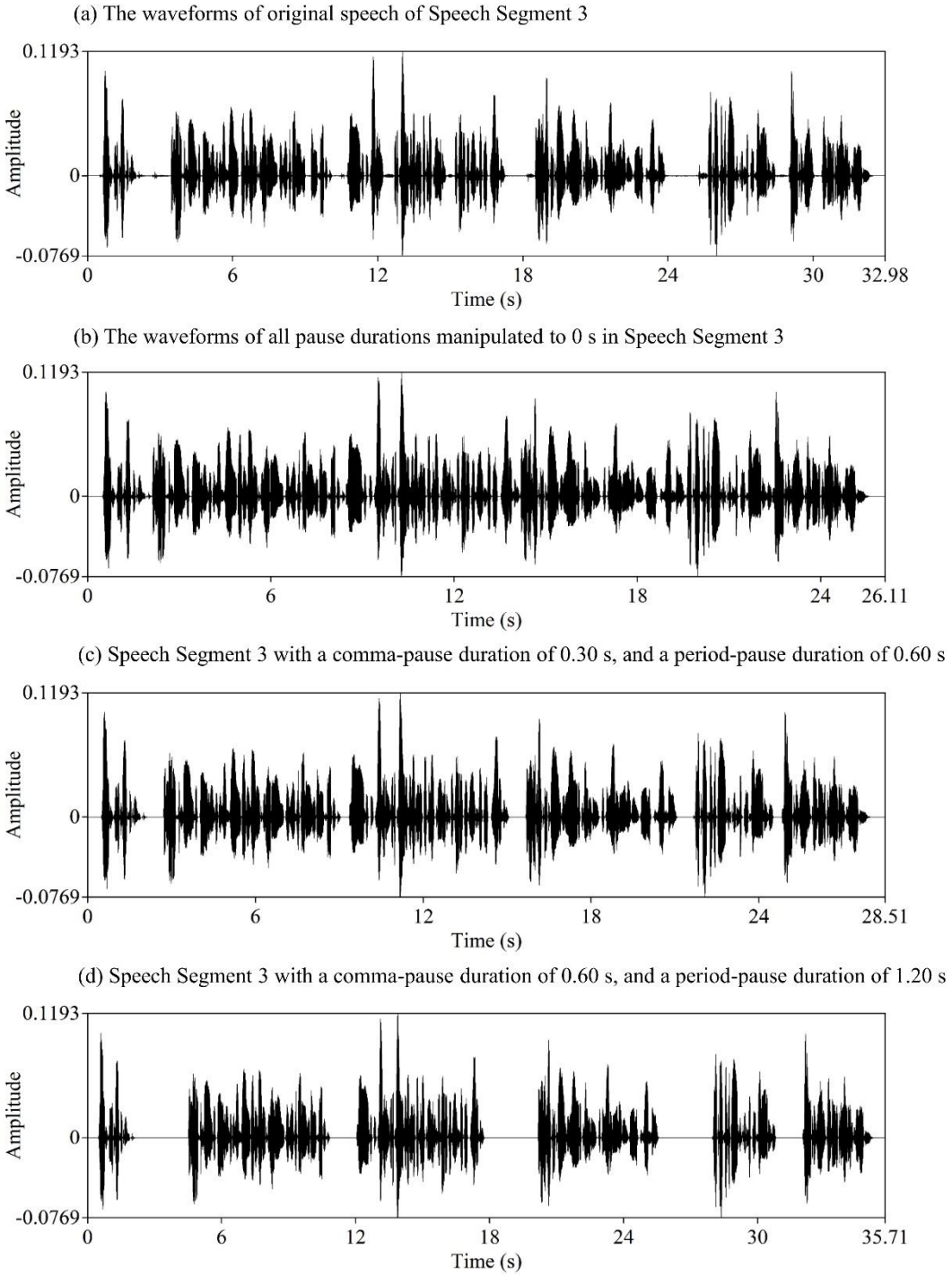


Figure 3.1 Examples of waveforms of the stimuli used throughout this thesis. (a) The waveforms of original speech of Speech Segment 3. (b) The waveforms of all pause durations manipulated to 0 s in Speech Segment 3. (c) Speech Segment 3 with a comma-pause duration of 0.30 s, and a period-pause duration of 0.60 s. (d) Speech Segment 3 with a comma-pause duration of 0.60 s, and a period-pause duration of 1.20 s. (Taken from "Praat" software window.)

3.2.3 Procedure

The experiment was conducted in two different places. The data from the native-English participants were obtained in Galway, Republic of Ireland, and in Fukuoka, Japan. The data from the Chinese and Japanese participants were obtained in Fukuoka, Japan. The procedure was the same as in Experiment 1, except that, in this experiment the participants rated the stimuli in three sessions on 12 evaluation items (**Table 2.1**). These 12 items were taken up from the 23 items used in Experiment 1. Because the comma- and period-pause durations were varied independently in the present experiment, fewer items were used to limit the total task duration. The first session was a short practice session. In the practice session, Speech Segments 3 and 4 (**Table 2.1**) were presented, each with comma- and period-pause duration of 0.6 s. These stimuli were the same for all participants, and the data were not used for further analysis. After the practice session was completed, two experimental sessions were carried out. In each session, 28 stimuli were randomly presented. The first stimulus and the last stimulus were the same, but the results of the first were not analyzed.

The stimuli were diotically presented to the participants through headphones 0.5 s after the participant pressed the "PLAY" button on the interface. When stimulus presentation was finished, the participants evaluated the stimulus on the 12 evaluation items, using pen and paper on which the 10-point rating scales were indicated. There was no time limit for participants to give each rating; the experiment was self-paced and took about 75 minutes, approximately. One limitation of Experiment 1 was also that the English proficiency of the non-native participants was checked only by asking whether they had actually performed an English proficiency test. In order to ascertain the English proficiency of the non-native participants, additional English listening and grammar tests were conducted

after the last session. That is, the participants were asked to write down the contents of the two speech stimuli used in the experiment, i.e., the spoken content of the female speaker (Speech Segment 3) and the male speaker (Speech Segment 4), as well as 5 randomly-selected sentences, each uttered by a different speaker, from an English-speech database consisting of short sentences (NTT-AT Multi-lingual speech database, 2002). To test English grammar knowledge, previous English-proficiency questions of the entrance exam of Kyushu University, Fukuoka, Japan, were used as well. All the participants (both Chinese and Japanese participants' groups) could answer at least 70 % of all the English questions (including the listening and grammar part). From this we assumed they had sufficient English capacity to participate in this listening experiment.

3.3 Results

3.3.1 Average evaluation scores for the 12 item

First of all, the distributions of the original data obtained from Experiment 2 were directly observed. Similar to Experiment 1, according to the average evaluation scores of each evaluation item, three types of distribution were observed: "Curve type" (see **Figure 3.2**), which had relatively low evaluation scores with short or long comma- and period-pause durations, but relatively high evaluation scores for intermediate pause durations; in this type of distribution, speech without any pauses got relatively low evaluation scores, but the original speech got relatively high evaluation scores. An approximately "steady decrease type" (see **Figure 3.3**), in which the evaluation scores decreased steadily as both comma- and period-pause duration became longer; speech without any pauses got the highest evaluation scores, but median or relatively low

evaluation scores for original speech. “No obvious change type” (see **Figure 3.4**), in which the evaluation scores did not change obviously as either comma- or period-pause duration changed; Neither speech without any pauses nor original speech showed obvious unevenness. Interestingly, the same tendencies were observed in the Chinese language group (see **Figures 3.5**), and in the Japanese language group (see **Figures 3.6**). From this, it could be expected that according to the type of evaluation items, pause duration affected the perception of the speech differently. (Average evaluation scores for each evaluation item are shown in **Appendix J**). Thus, factor analysis was conducted on the data of Experiment 2.

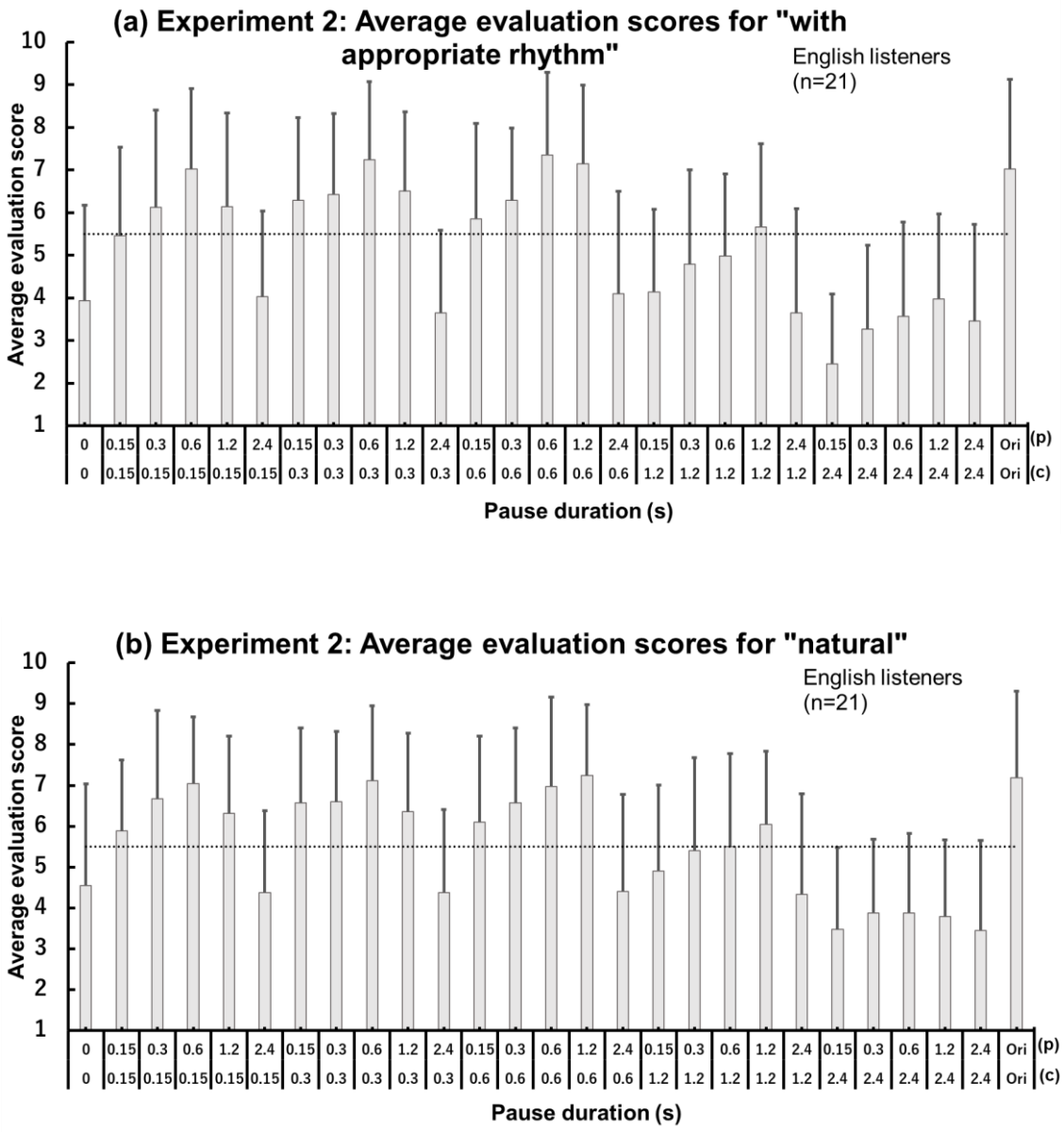


Figure 3.2 Examples of a "curve type" distribution in the English group. (a) Average evaluation scores for the item "with appropriate rhythm". (b) Average evaluation scores for "natural". The error bars show the standard deviation. The black dotted line shows the scale midpoint. ''ori'' indicates the original speech. ''(p)'' in the upper x-axis indicates the pause duration for periods. ''(c)'' in the lower x-axis indicates the pause duration for commas.

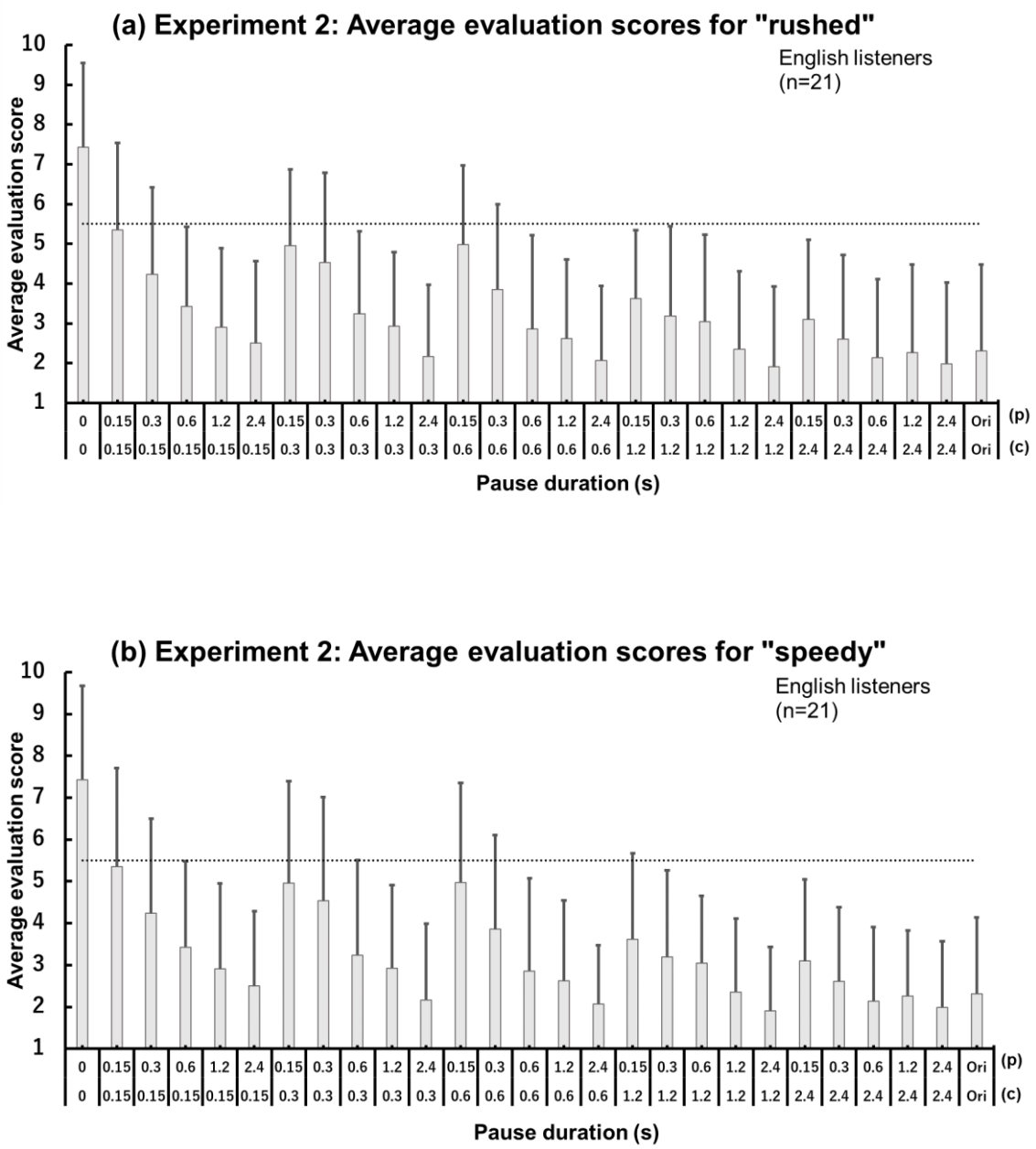


Figure 3.3 Examples of a "steady decrease type" distribution in the English group. (a) Average evaluation scores for the item "rushed". (b) Average evaluation scores for "speedy". The error bars show standard deviations. The black dotted line shows the scale midpoint. "ori" indicates the original speech. "(p)" in the upper x-axis indicates the pause duration for periods. "(c)" in the lower x-axis indicates the pause duration for commas.

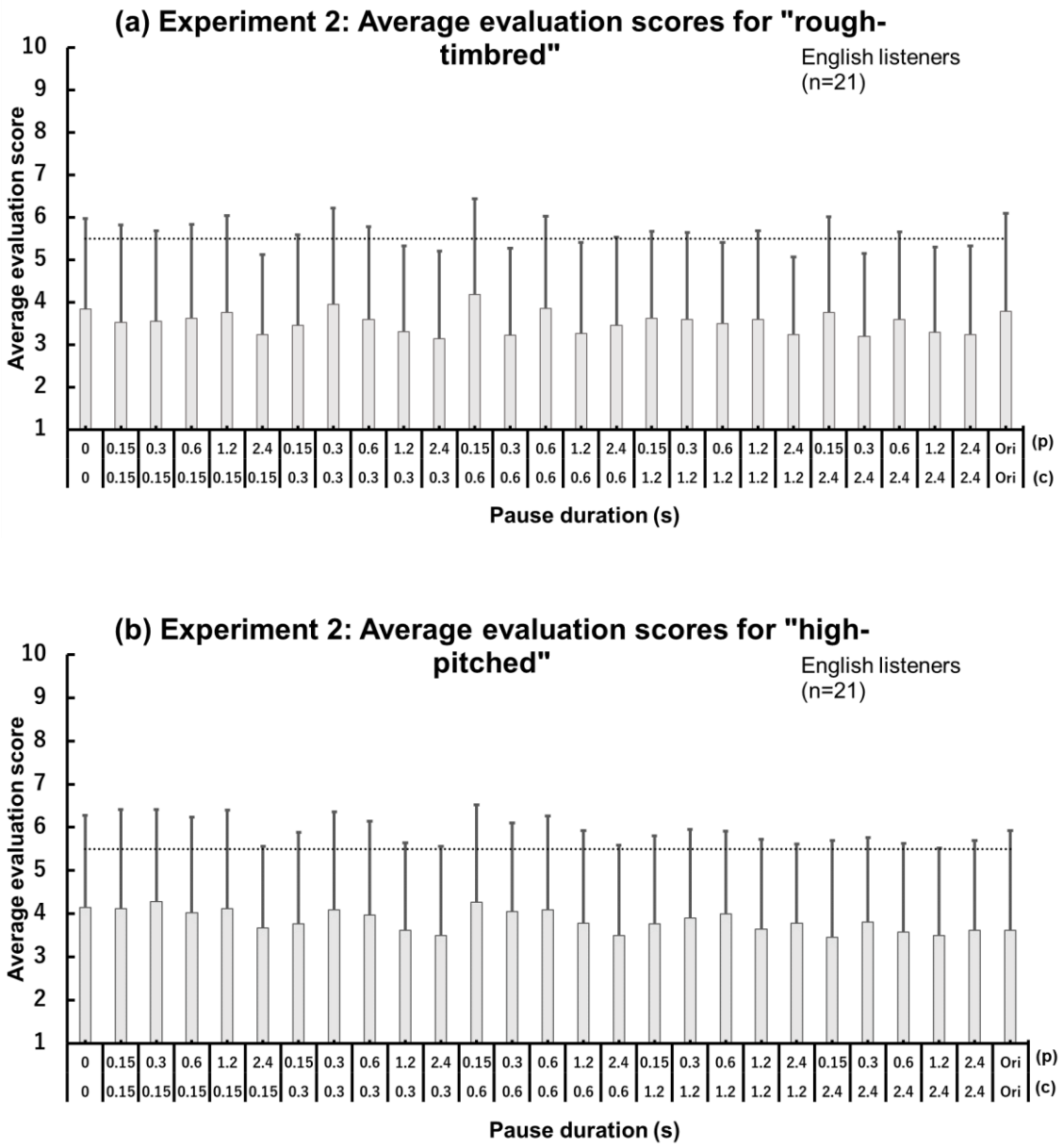
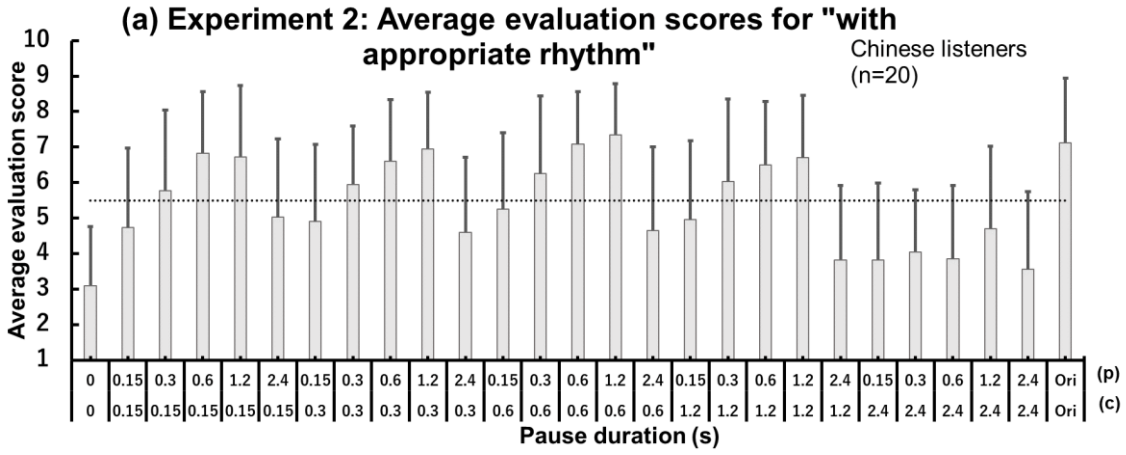
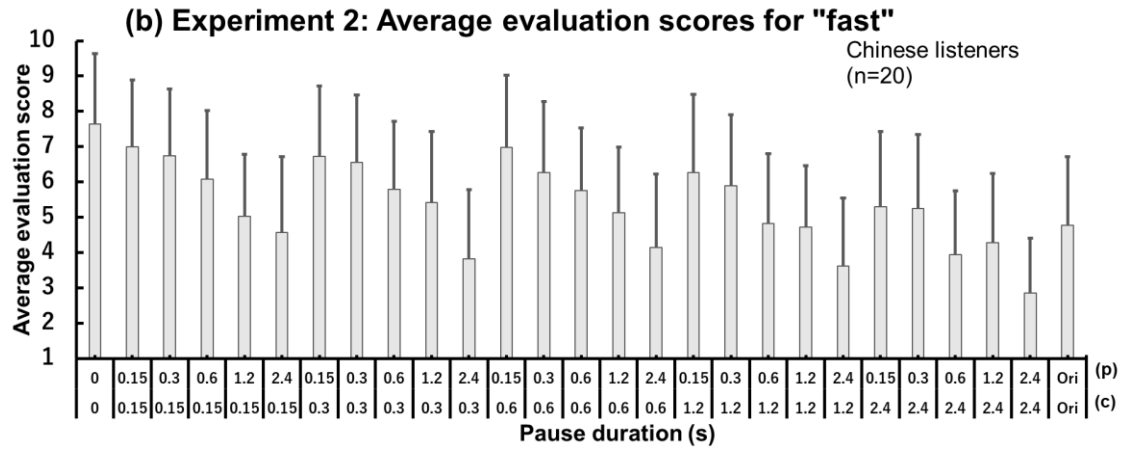


Figure 3.4 Examples of a "no obvious type" distribution in the English group. (a) Average evaluation scores for the item "rough-timbred". (b) Average evaluation scores for "high-pitched". The error bars show standard deviations. The black dotted line shows the scale midpoint. ''ori'' indicates the original speech. ''(p)'' in the upper x-axis indicates the pause duration for periods. ''(c)'' in the lower x-axis indicates the pause duration for commas.


(a) Experiment 2: Average evaluation scores for "with appropriate rhythm"
Chinese listeners (n=20)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
0 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 Ori (p)
0 0.15 0.15 0.15 0.15 0.15 0.3 0.3 0.3 0.3 0.3 0.6 0.6 0.6 0.6 0.6 1.2 1.2 1.2 1.2 1.2 2.4 2.4 2.4 2.4 2.4 Ori (c)
Pause duration (s)


(b) Experiment 2: Average evaluation scores for "fast"
Chinese listeners (n=20)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
0 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 Ori (p)
0 0.15 0.15 0.15 0.15 0.15 0.3 0.3 0.3 0.3 0.3 0.6 0.6 0.6 0.6 0.6 1.2 1.2 1.2 1.2 1.2 2.4 2.4 2.4 2.4 2.4 Ori (c)
Pause duration (s)

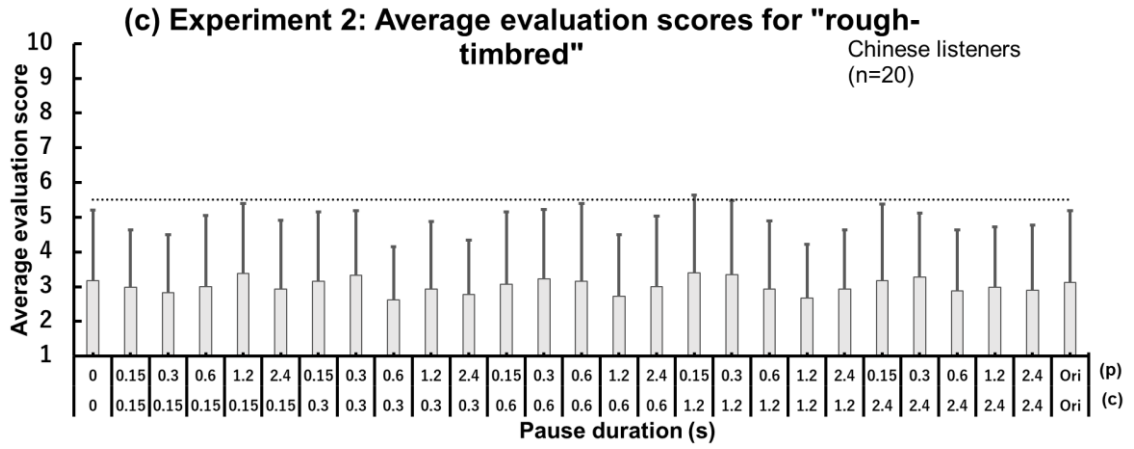
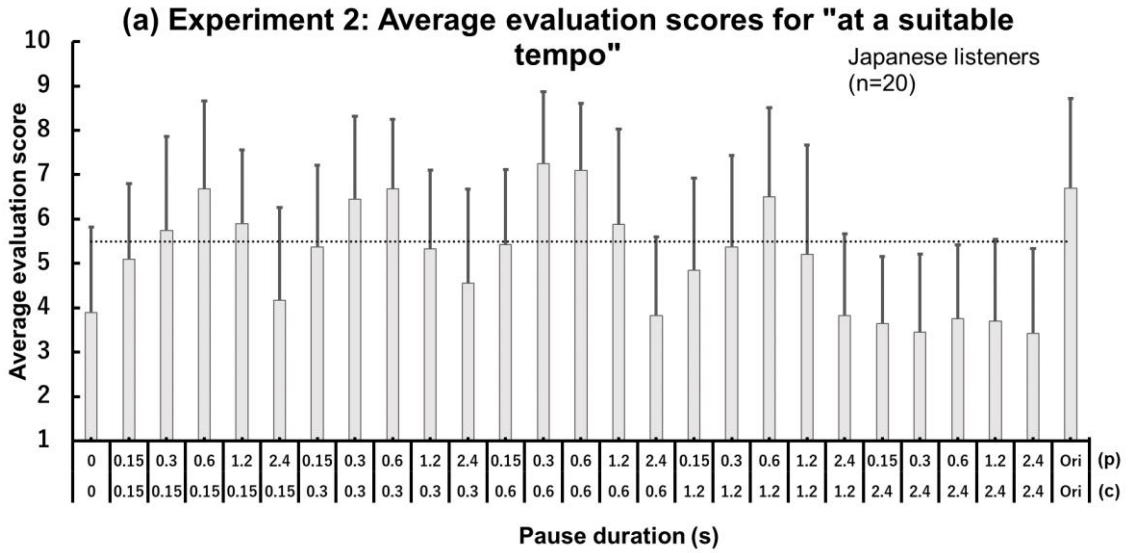
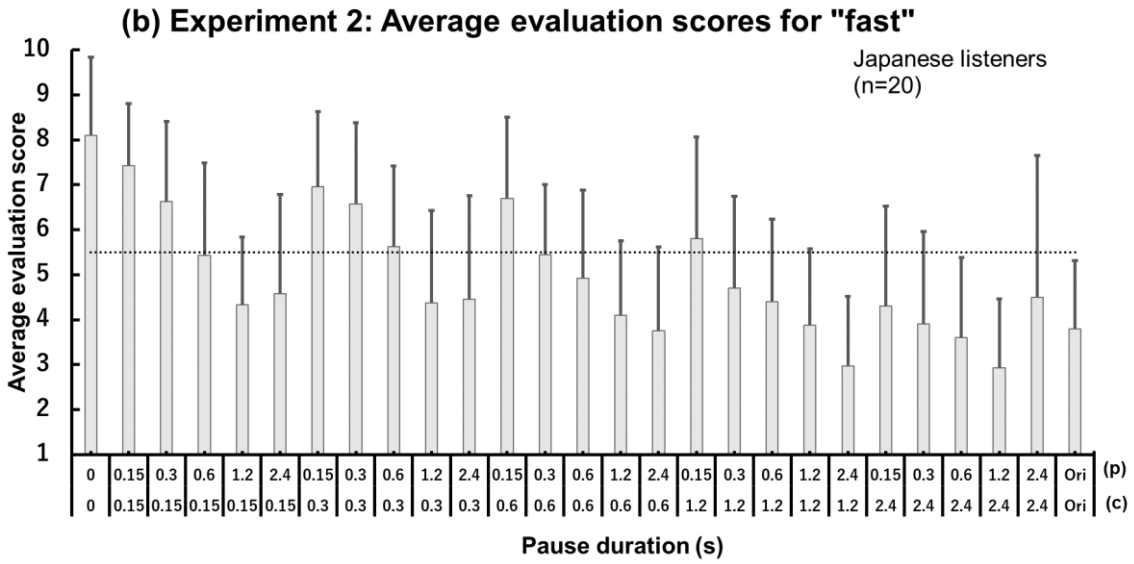


Figure 3.5 Examples of three types of distributions in the Chinese group. (a) "Curve type" distribution: Average evaluation scores for the item "with appropriate rhythm". (b) "Steady decrease type" distribution: Average evaluation scores for "fast". (c) "No obvious type" distribution: Average evaluation scores for "rough-timbred". The error bars show standard deviations. The black dotted line shows the scale midpoint. ''ori'' indicates the original speech. ''(p)'' in the upper x-axis indicates the pause duration for periods. ''(c)'' in the lower x-axis indicates the pause duration for commas.





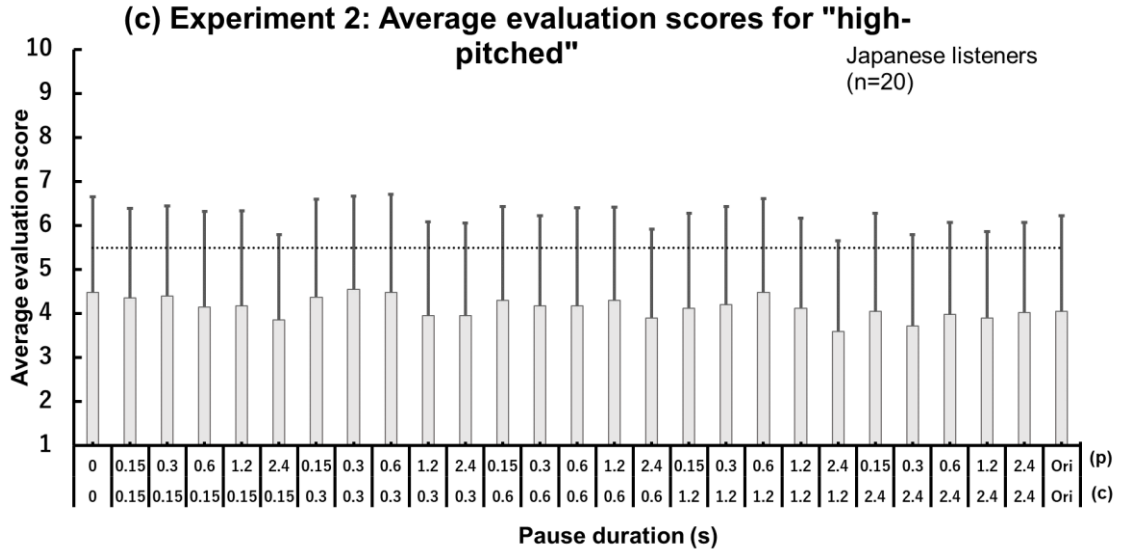


Figure 3.6 Examples of three types of distributions in the Japanese group. (a) “Curve type” distribution: Average evaluation scores for the item “at a suitable tempo”. (b) “Steady decrease type” distribution: Average evaluation scores for “fast”. (c) “No obvious type” distribution: Average evaluation scores for “high-pitched”. The error bars show standard deviations. The black dotted line shows the scale midpoint. ‘‘ori’’ indicates the original speech. ‘‘(p)’’ in the upper x-axis indicates the pause duration for periods. ‘‘(c)’’ in the lower x-axis indicates the pause duration for commas.

3.3.2 Results of factor analysis

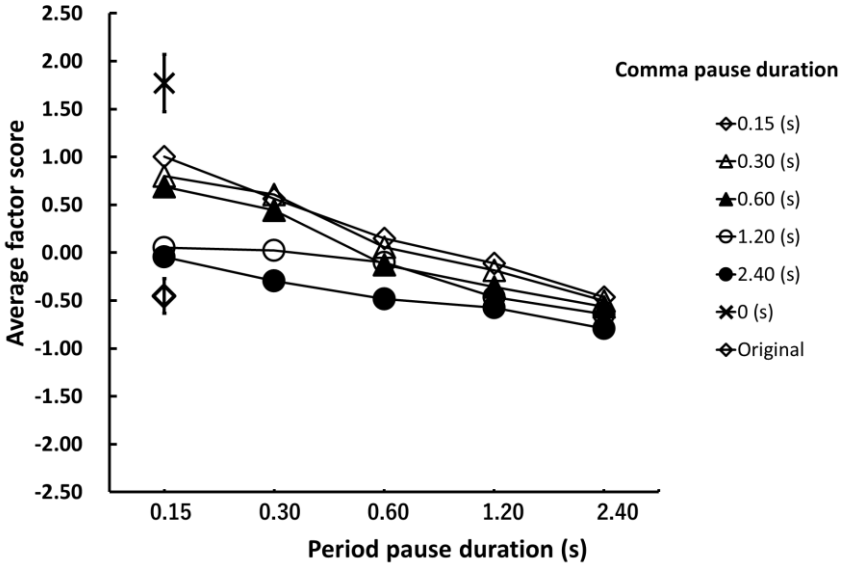
The results were analyzed using the same protocol as used in Experiment 1. Since three of the native-English participants did not evaluate the stimuli on three or more evaluation items, their data were not analyzed. Five native-English participants had missed one evaluation item, and 7 had provided no score on two evaluation items. Their data were nevertheless included in the PCA; instead of the blank data entry we added the median score of the rating scale (5.5). Before performing PCA, KMO-tests showed that the data sampling was adequate overall for the native-English participants (0.852), the Chinese participants (0.877), and the Japanese participants (0.914). [Bartlett's tests of sphericity were also all significant ($p<0.001$)].

For the native-English and the Chinese language group, three factors were extracted from PCA with varimax rotation, and two factors for the Japanese language group. Similar to the results of Experiment 1, for all three language groups, the first factor could be interpreted as the Speech Naturalness factor and the second factor as the Speech Rate factor. The Speech Naturalness factor included the evaluation items “with appropriate rhythm”, “at a suitable tempo”, “natural”, “with appropriate pause duration”, “skillful”, “well-practiced”, and “friendly”. The Speech Rate factor included “speedy”, “rushed”, and “fast” for all three language groups. Also similar to Experiment 1, the third factor that appeared in the PCA for the native-English and the Chinese language group related to sound quality. The cumulative percentages of variance for the Speech Naturalness and the Speech Rate factor were over 60% for all three language groups. The cumulative percentage of variance at the third factor in the English and Chinese language group reached 71%. Only the Speech Naturalness and the Speech Rate factor were discussed here.

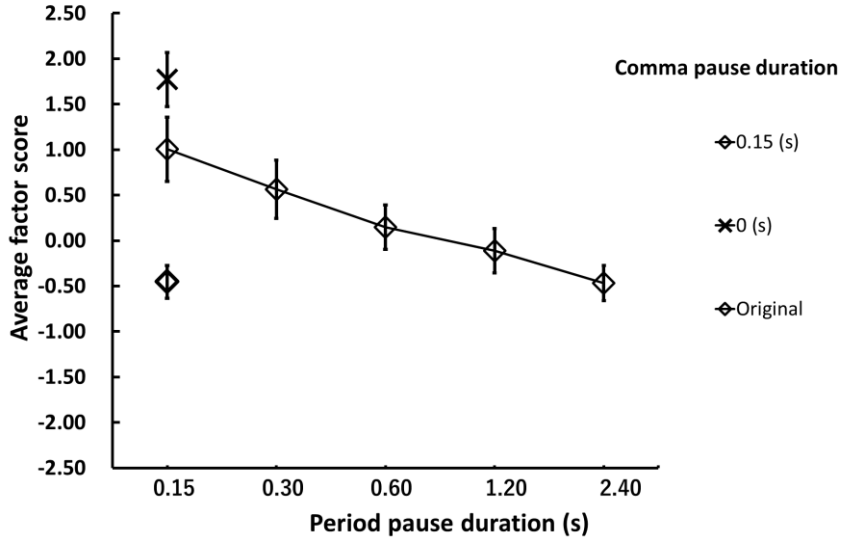
3.3.3 Results of Speech Rate factor

Because Shapiro-Wilk tests showed that the Speech Rate factor scores were not normally distributed for all three language groups, they were analyzed as in Experiment 1. The Friedman tests were significant [native-English group ($\chi 2$ (df= 26, n= 21) = 281.6, $p < 0.0001$; Chinese group ($\chi 2$ (df= 26, n= 20) = 323.5, $p < 0.0001$; Japanese group ($\chi 2$ (df= 26, n= 20) = 395.5, $p < 0.0001$], and the Kendall's Coefficient of Concordance test showed that the factor scores obtained were highly similar among the groups (Kendall's W= 0.98, $p < 0.01$, n= 3, k= 27). The factor scores for the original speech were as expected based on the physical durations of the speech stimuli as shown in **Figures 3.7-3.9**. Based on the 95%-confidence intervals in **Figures 3.7-3.9**, for all three language groups, stimuli with a period-pause duration below 1.2 s and a comma-pause duration below 0.6 s had significantly higher factor scores than the original speech. Similar to Experiment 1, the results clearly show that when the comma- and period-pause duration became longer, the listeners in all three language groups perceived a slower speech rate, even though only the pause duration was adjusted and not the speech itself. The average Speech Rate factor scores showed a steady decrease as pause duration increased from 0.15 to 2.4 s for all three groups. The average factor score for the 0-s condition was, as expected, the highest. Overall, Experiment 2 confirmed that the perceived speech rate thus can be influenced by only manipulating pause duration.

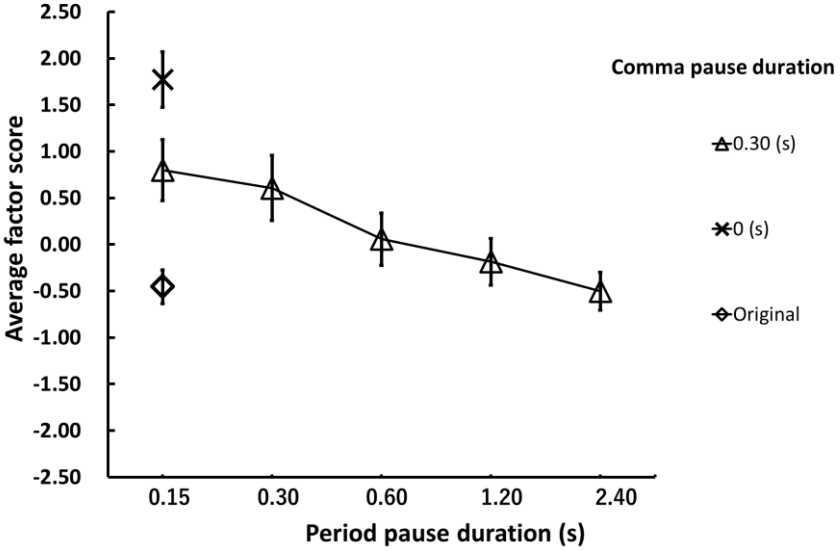
(a) Experiment 2: Average factor scores for the Speech Rate factor from English listeners (n=21)



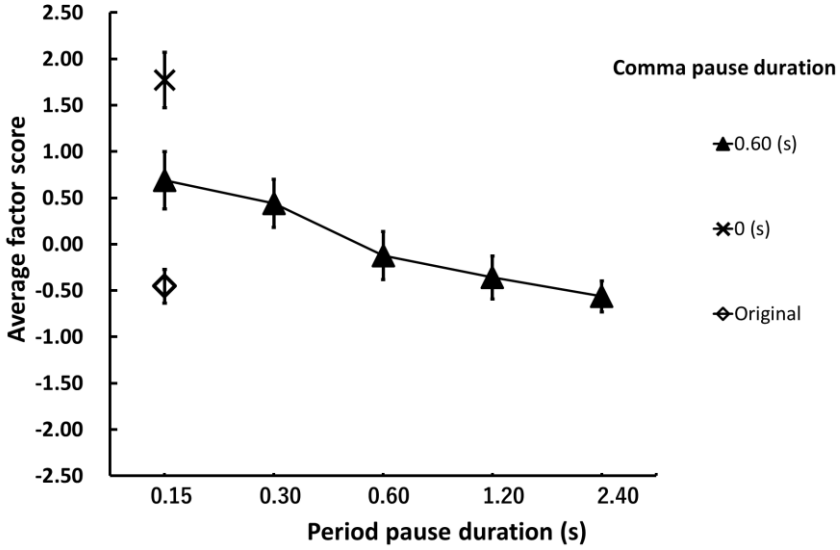
(b) Experiment 2: Average factor scores for the Speech Rate factor from English listeners (n=21)

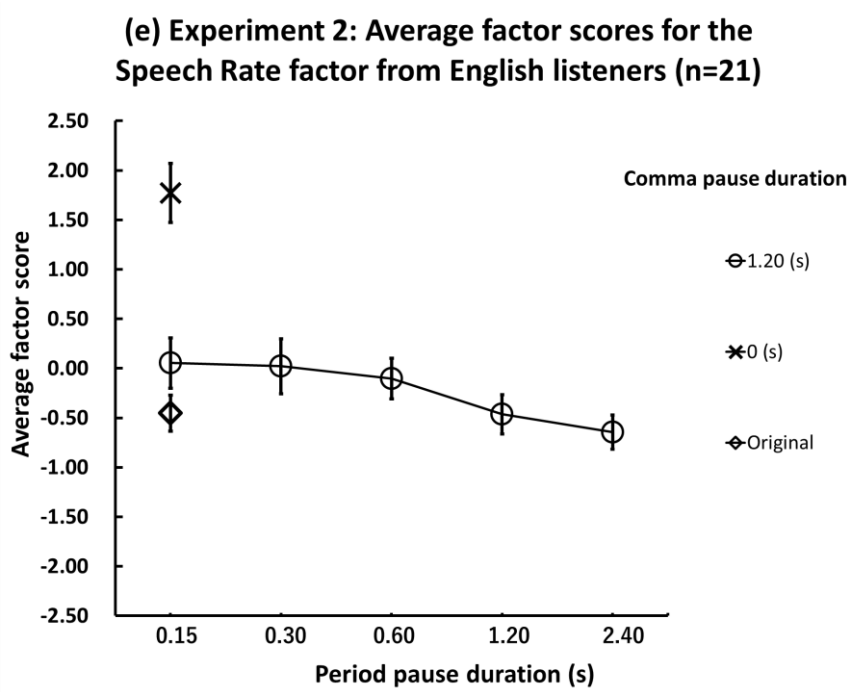


(c) Experiment 2: Average factor scores for the Speech Rate factor from English listeners (n=21)



(d) Experiment 2: Average factor scores for the Speech Rate factor from English listeners (n=21)





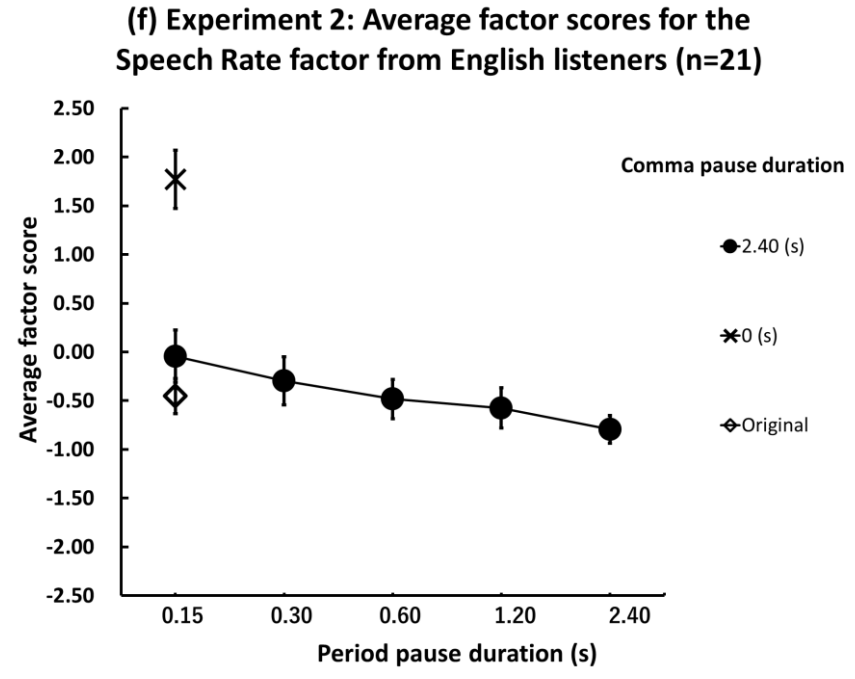
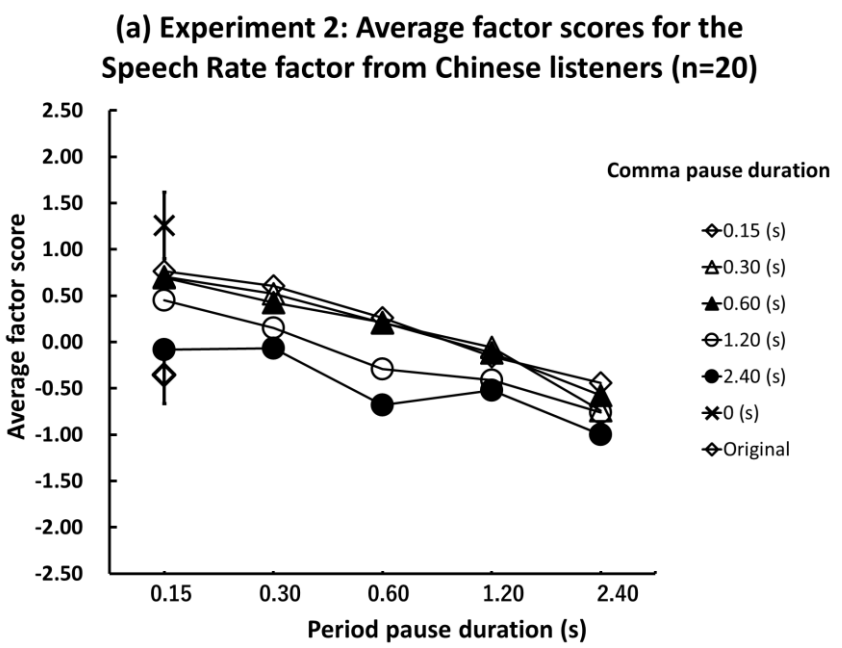
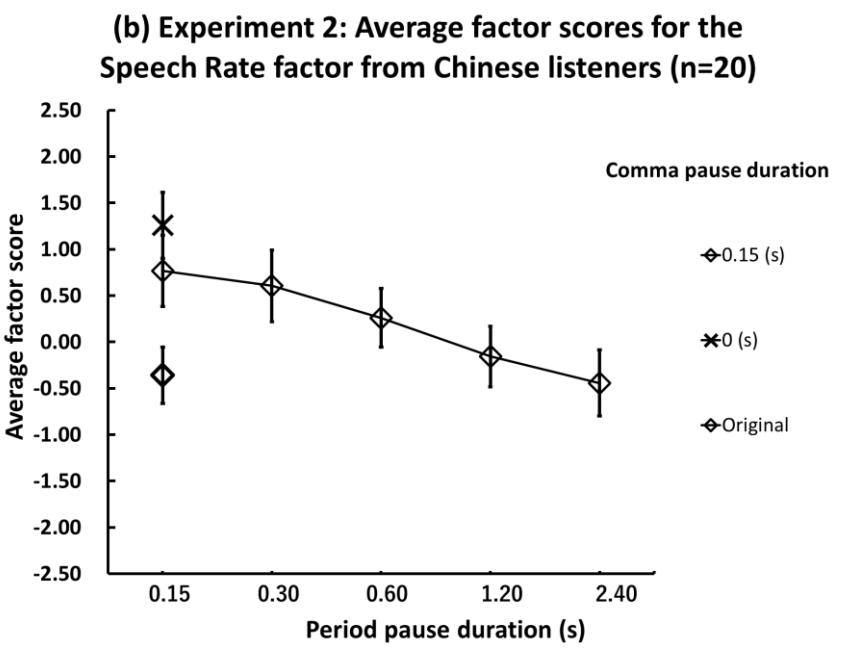
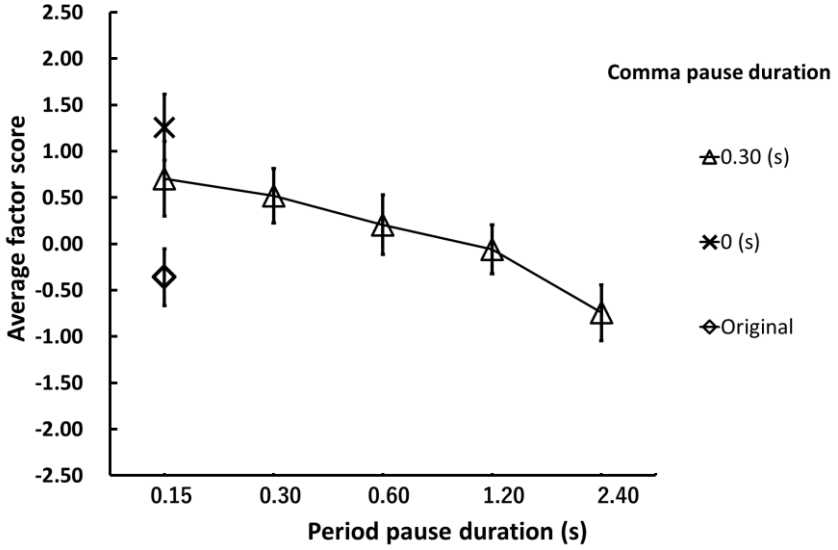


Figure 3.7 Results of Experiment 2. (a) The average factor scores for the Speech Rate factor from the English language group. (b) – (f) The average factor scores for the Speech Rate factor for comma-pause durations of 0.15 s, 0.3 s, 0.6 s 1.2 s, and 2.4 s, respectively. The error bars show the 95%-confidence intervals.

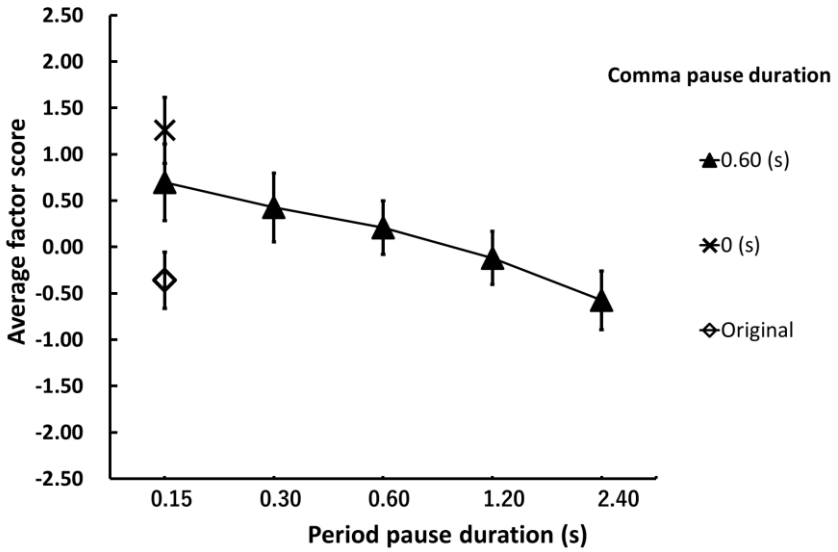

(a) Experiment 2: Average factor scores for the Speech Rate factor from Chinese listeners (n=20)
Average factor score
2.50
2.00
1.50
1.00
0.50
0.00
-0.50
-1.00
-1.50
-2.00
-2.50
0.15
0.30
0.60
1.20
2.40
Period pause duration (s)
Comma pause duration
0.15 (s)
0.30 (s)
0.60 (s)
1.20 (s)
2.40 (s)
0 (s)
Original

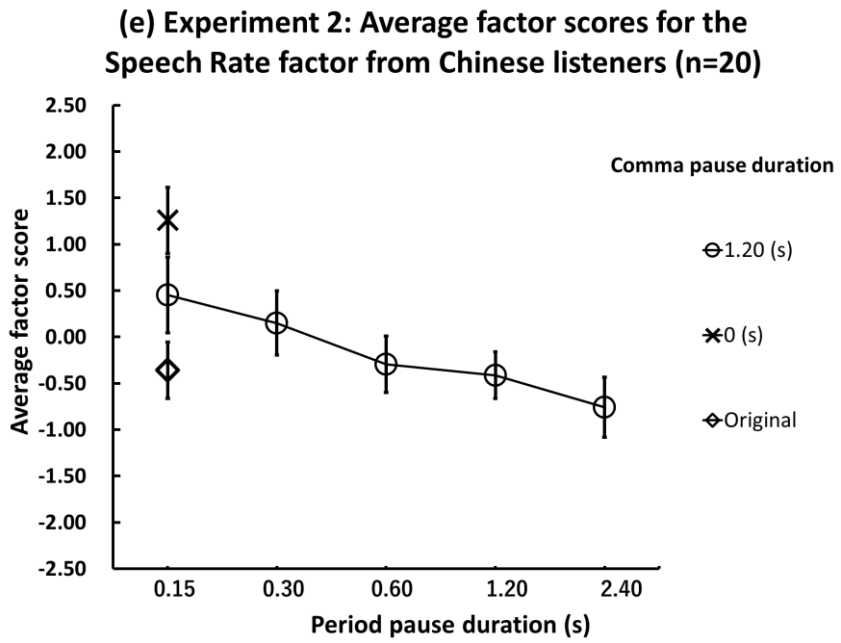

(b) Experiment 2: Average factor scores for the Speech Rate factor from Chinese listeners (n=20)
Average factor score
2.50
2.00
1.50
1.00
0.50
0.00
-0.50
-1.00
-1.50
-2.00
-2.50
0.15
0.30
0.60
1.20
2.40
Period pause duration (s)
Comma pause duration
0.15 (s)
0 (s)
Original

(c) Experiment 2: Average factor scores for the Speech Rate factor from Chinese listeners (n=20)



(d) Experiment 2: Average factor scores for the Speech Rate factor from Chinese listeners (n=20)





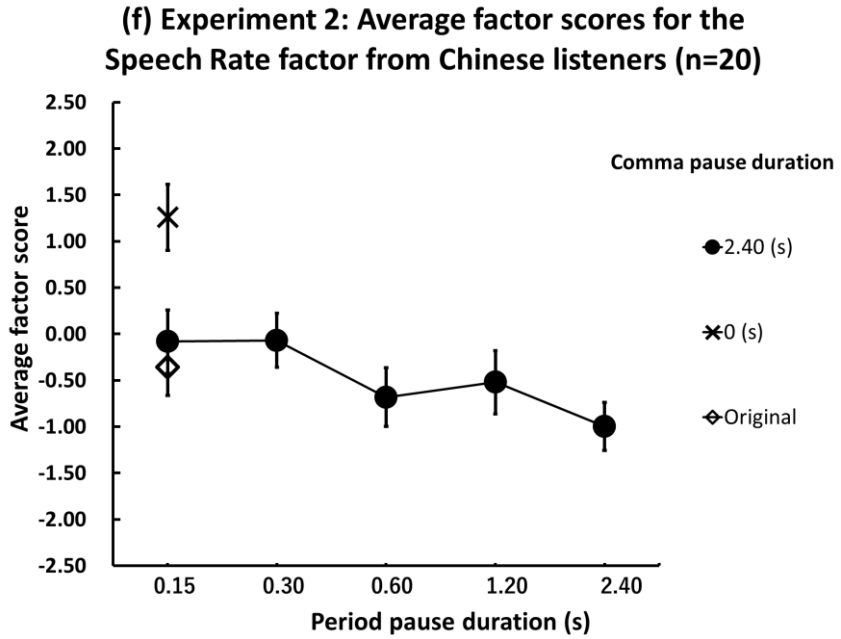
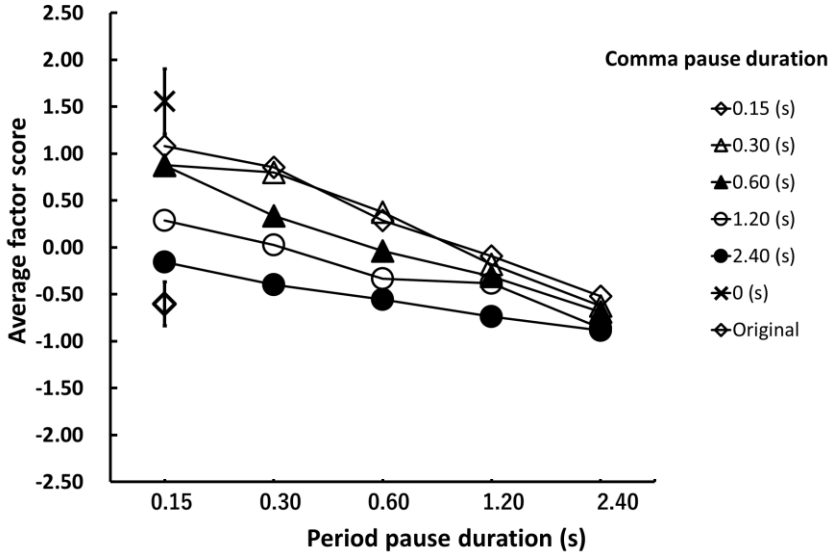
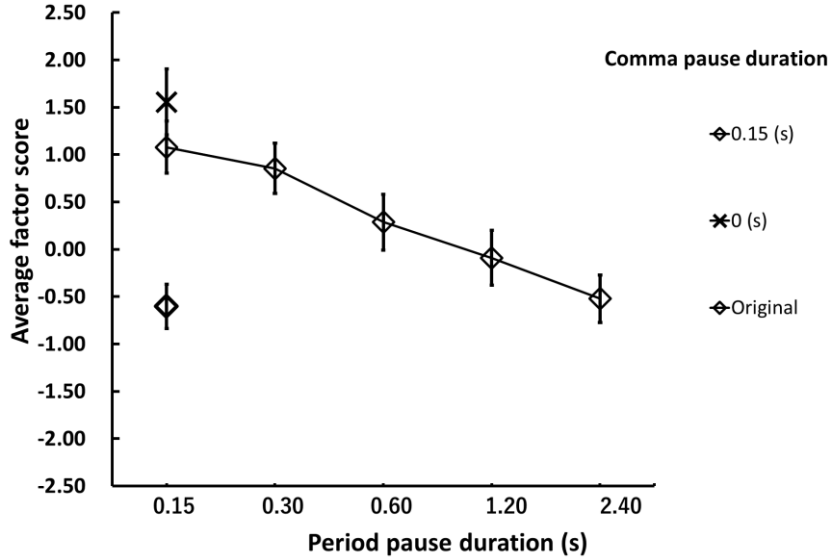


Figure 3.8 Results of Experiment 2. (a) The average factor scores for the Speech Rate factor from the Chinese language group. (b) – (f) The average factor scores for the Speech Rate factor for comma-pause durations of 0.15 s, 0.3 s, 0.6 s 1.2 s, and 2.4 s, respectively. The error bars show the 95%-confidence intervals.

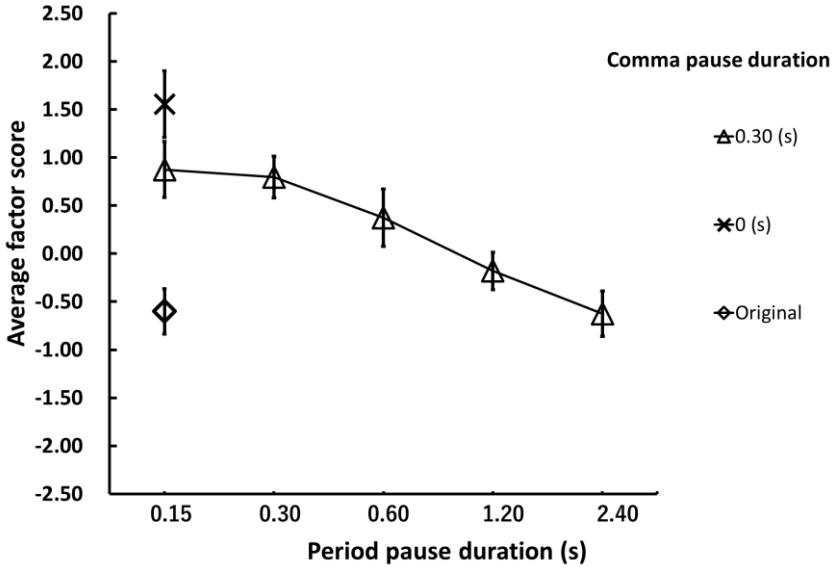
(a) Experiment 2: Average factor scores for the Speech Rate factor from Japanese listeners (n=20)



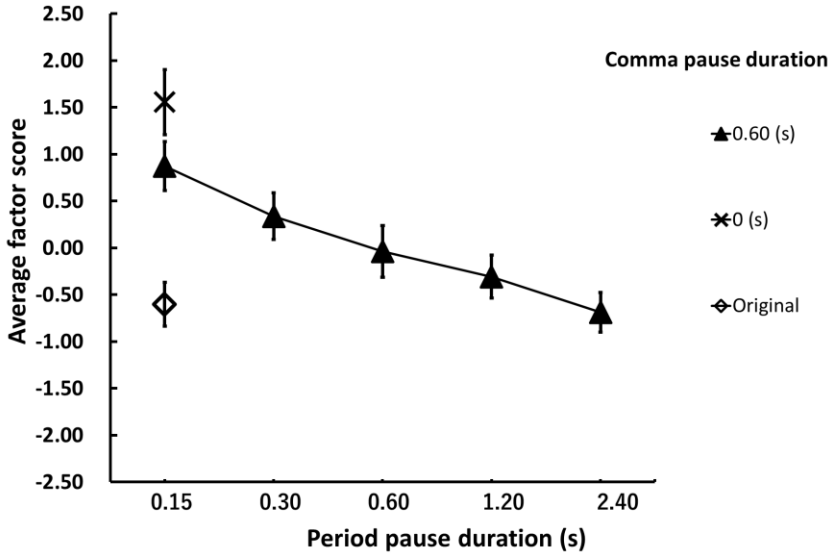
(b) Experiment 2: Average factor scores for the Speech Rate factor from Japanese listeners (n=20)

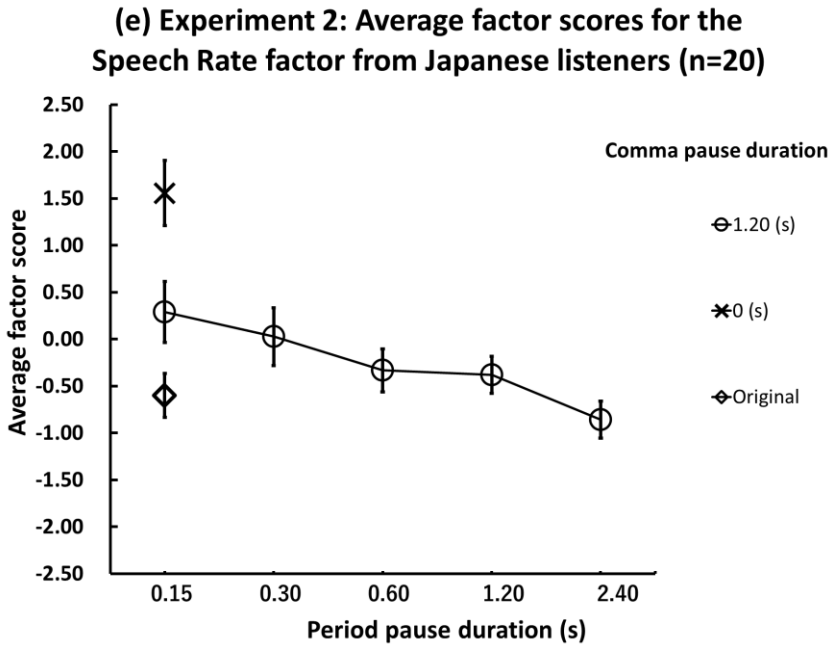


(c) Experiment 2: Average factor scores for the Speech Rate factor from Japanese listeners (n=20)



(d) Experiment 2: Average factor scores for the Speech Rate factor from Japanese listeners (n=20)





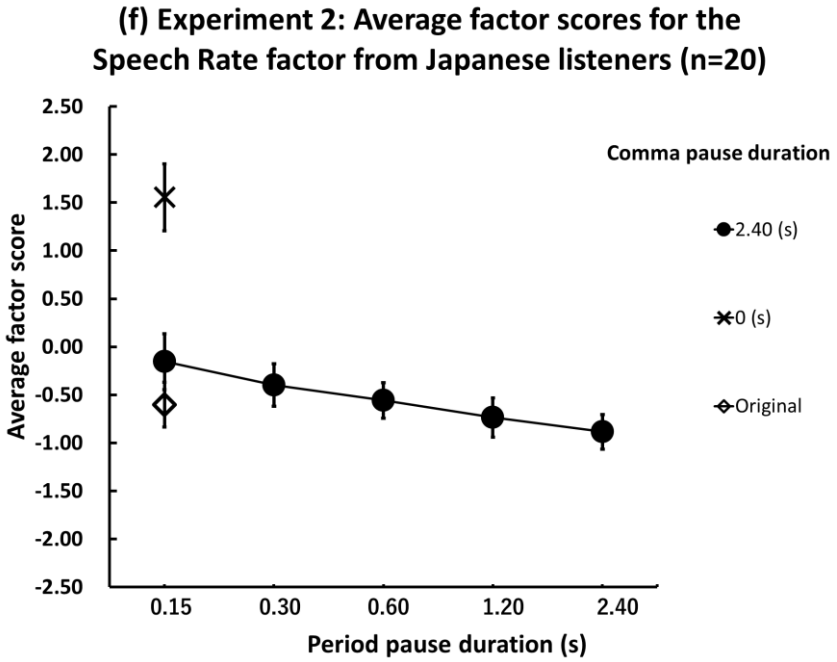
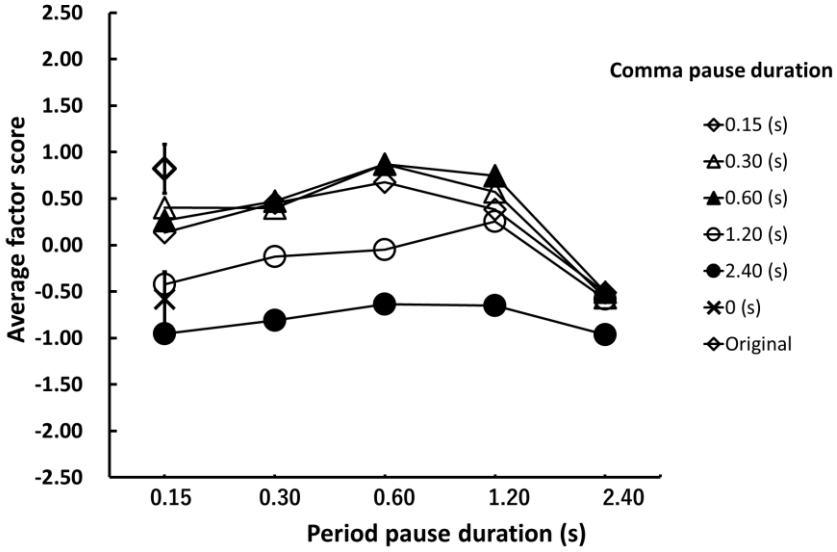


Figure 3.9 Results of Experiment 2. (a) The average factor scores for the Speech Rate factor from the Japanese language group. (b) – (f) The average factor scores for the Speech Rate factor for comma-pause durations of 0.15 s, 0.3 s, 0.6 s 1.2 s, and 2.4 s, respectively. The error bars show the 95%-confidence intervals.

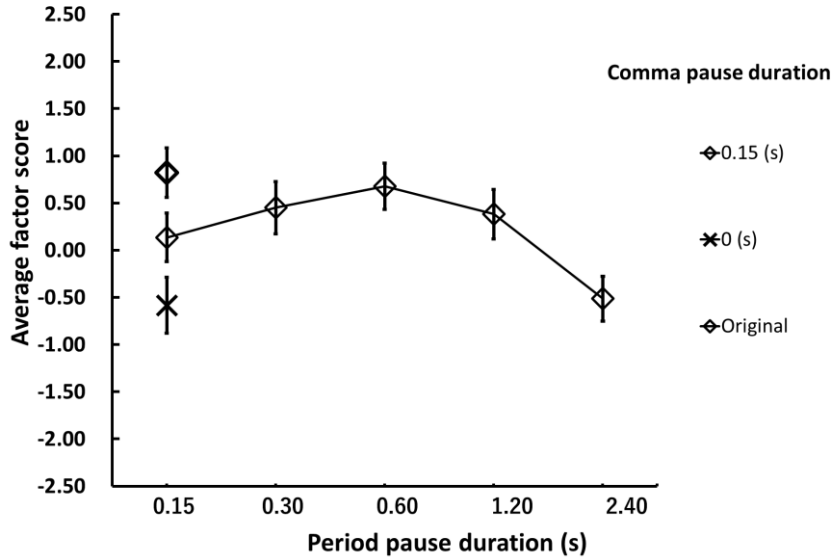
3.3.4 Results for the Naturalness factor

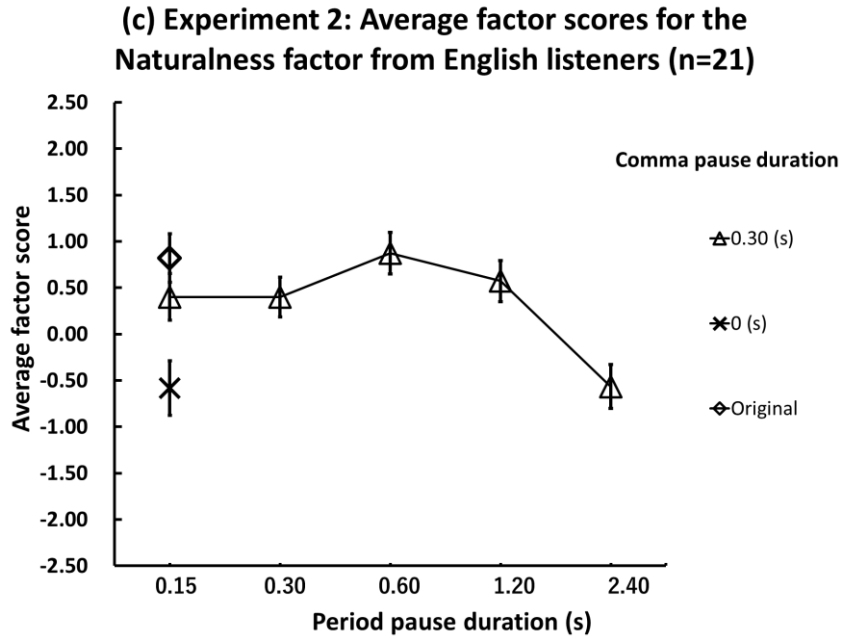
The Friedman tests over the factor scores of the Naturalness factor were significant [native-English group ($\chi 2$ (df= 26, n= 21) = 328.7, $p<0.0001$; Chinese group ($\chi 2$ (df= 26, n= 20) = 289.3, $p<0.0001$; Japanese group ($\chi 2$ (df= 26, n= 20) = 336.2, $p<0.0001$], and highly similar for the three language groups (Kendall's W= 0.95, $p<0.01$, n= 3, k= 27), as can be seen in **Figures 3.10-3.12**. In the native-English group, the factor score for speech sentences with a comma-pause duration and a period-pause duration of 0.6 s was the highest (0.87), slightly above the factor score for sentences with a comma-pause of 0.3 s and a period-pause of 0.6 s (0.84). Remarkably, these factor scores were higher than that for the original speech (0.82). For the Chinese language group, the original speech got the highest factor score (0.81), closely followed by the speech sentences with a comma-pause duration of 0.6 s and a period-pause duration of 1.2 s (0.79). For the Japanese language group, the factor score for speech sentences with a comma- and period-pause duration of 0.6 s was the highest (1.20), again, surprisingly, exceeding the factor score obtained for the original speech (0.85). From the 95%-confidence intervals in Figures **3.10-3.12**, it can be seen that there were no significant differences between the factor scores for the original speech and the stimuli with adequately manipulated pause durations, mentioned above.

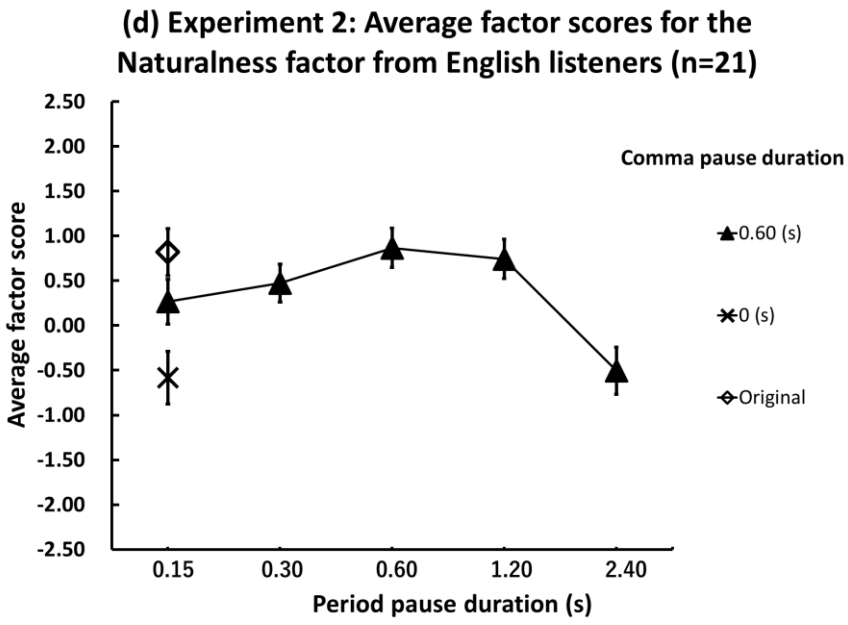
(a) Experiment 2: Average factor scores for the Naturalness factor from English listeners (n=21)

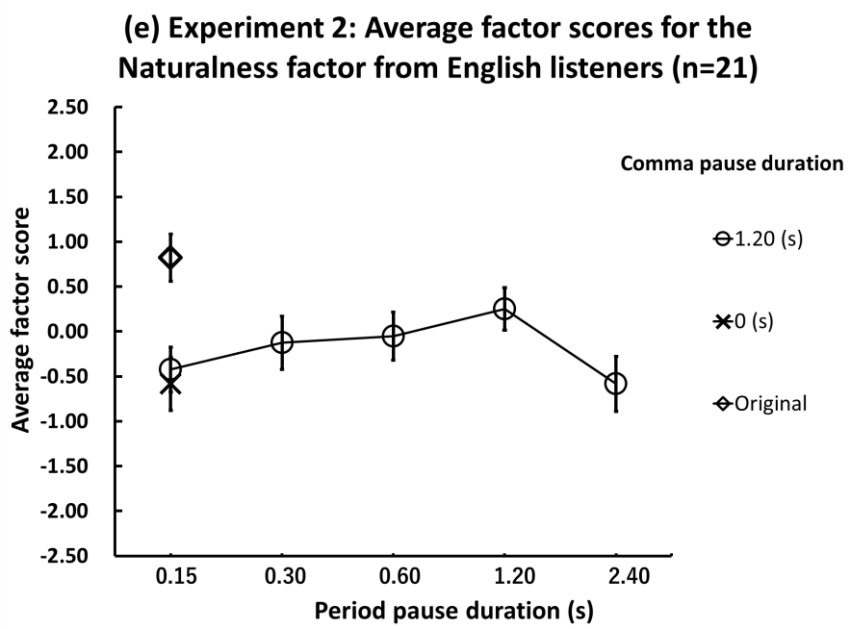


(b) Experiment 2: Average factor scores for the Naturalness factor from English listeners (n=21)




(c) Experiment 2: Average factor scores for the Naturalness factor from English listeners (n=21)
Average factor score
2.50
2.00
1.50
1.00
0.50
0.00
-0.50
-1.00
-1.50
-2.00
-2.50
0.15
0.30
0.60
1.20
2.40
Period pause duration (s)
Comma pause duration
0.30 (s)
0 (s)
Original


(d) Experiment 2: Average factor scores for the Naturalness factor from English listeners (n=21)
Average factor score
2.50
2.00
1.50
1.00
0.50
0.00
-0.50
-1.00
-1.50
-2.00
-2.50
0.15
0.30
0.60
1.20
2.40
Period pause duration (s)
Comma pause duration
0.60 (s)
0 (s)
Original



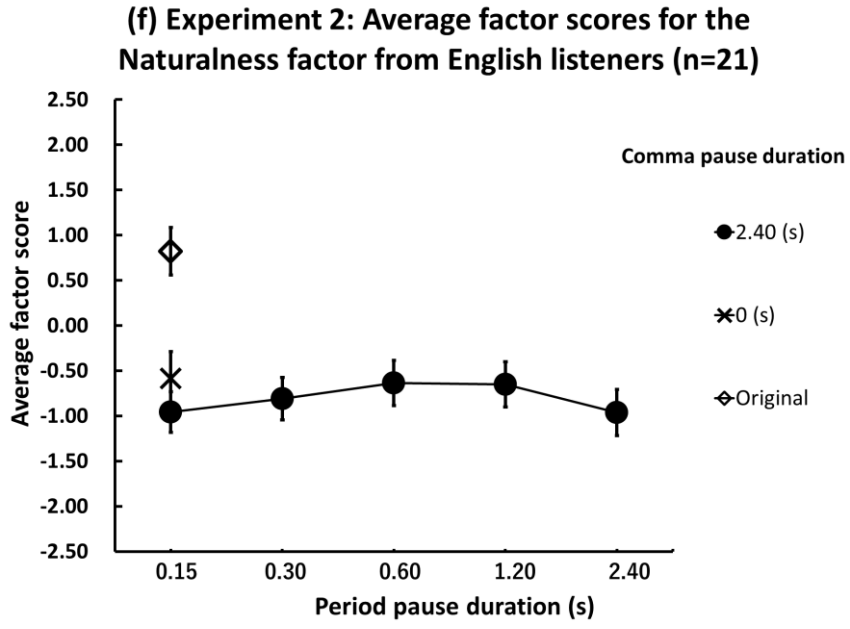
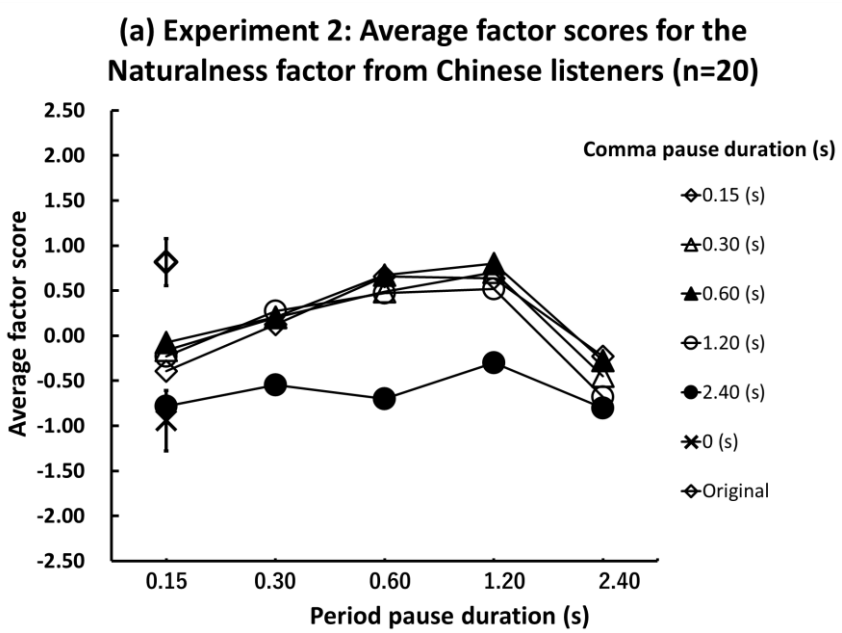
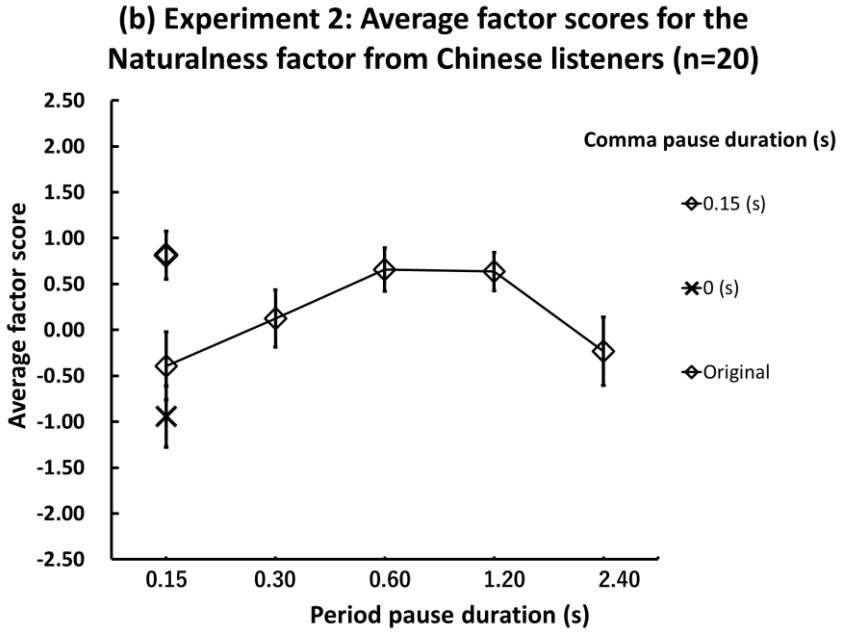
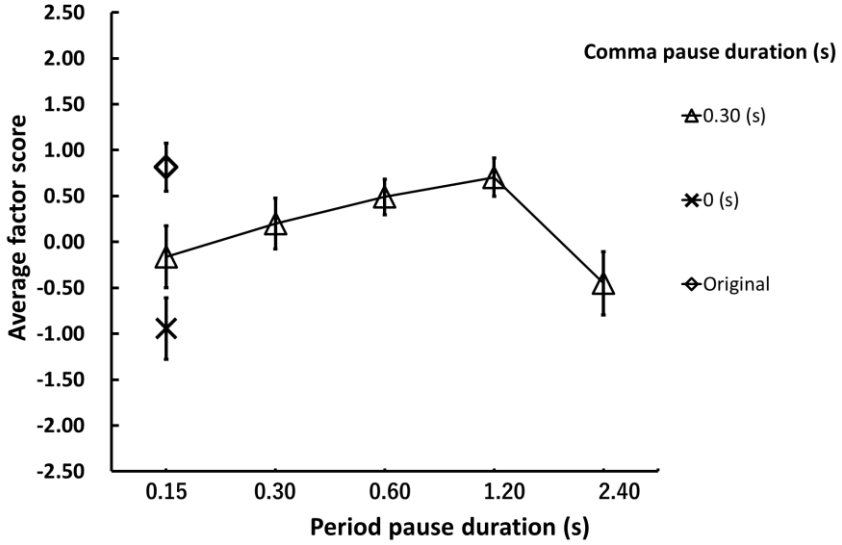


Figure 3.10 Results of Experiment 2. (a) The average factor scores for the Naturalness factor from the English language group. (b) – (f) The average factor scores for the Naturalness factor for comma-pause durations of 0.15 s, 0.3 s, 0.6 s 1.2 s, and 2.4 s, respectively. The error bars show the 95%-confidence intervals.

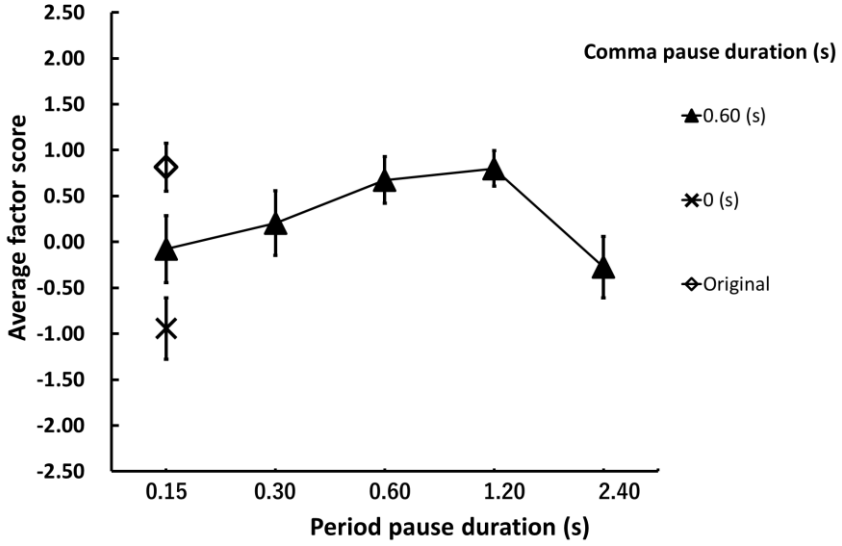

(a) Experiment 2: Average factor scores for the Naturalness factor from Chinese listeners (n=20)
Average factor score
2.50
2.00
1.50
1.00
0.50
0.00
-0.50
-1.00
-1.50
-2.00
-2.50
0.15
0.30
0.60
1.20
2.40
Period pause duration (s)
Comma pause duration (s)
0.15 (s)
0.30 (s)
0.60 (s)
1.20 (s)
2.40 (s)
0 (s)
Original

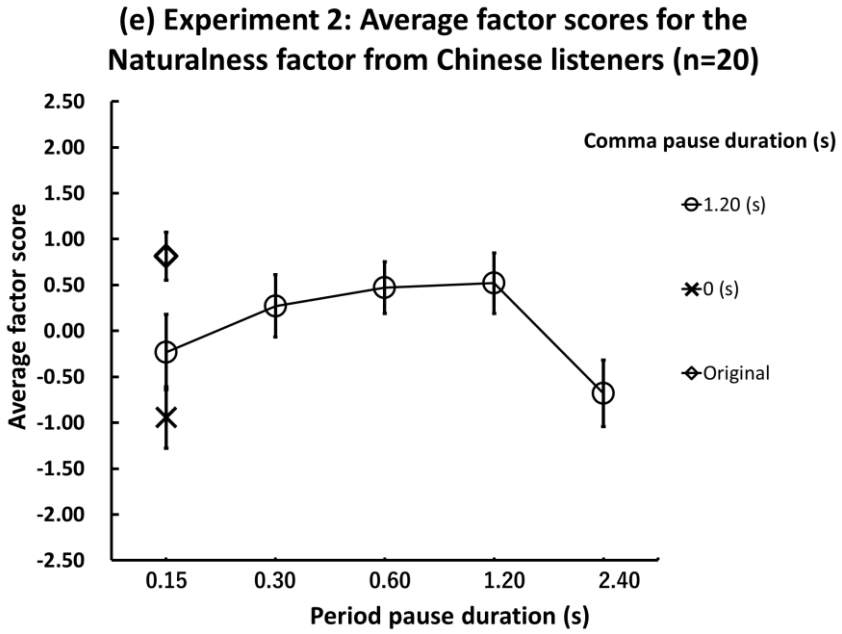

(b) Experiment 2: Average factor scores for the Naturalness factor from Chinese listeners (n=20)
Average factor score
2.50
2.00
1.50
1.00
0.50
0.00
-0.50
-1.00
-1.50
-2.00
-2.50
0.15
0.30
0.60
1.20
2.40
Period pause duration (s)
Comma pause duration (s)
0.15 (s)
0 (s)
Original

(c) Experiment 2: Average factor scores for the Naturalness factor from Chinese listeners (n=20)



(d) Experiment 2: Average factor scores for the Naturalness factor from Chinese listeners (n=20)





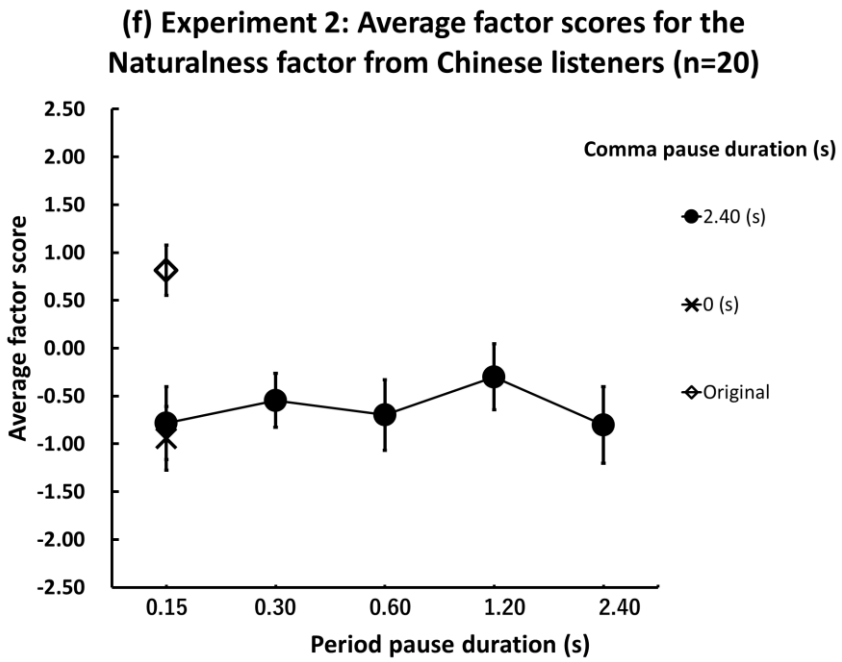
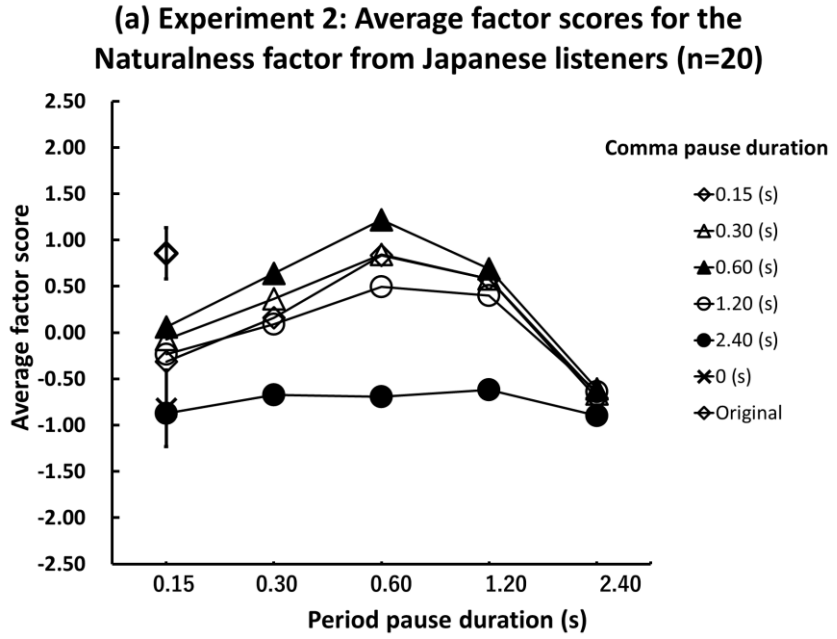
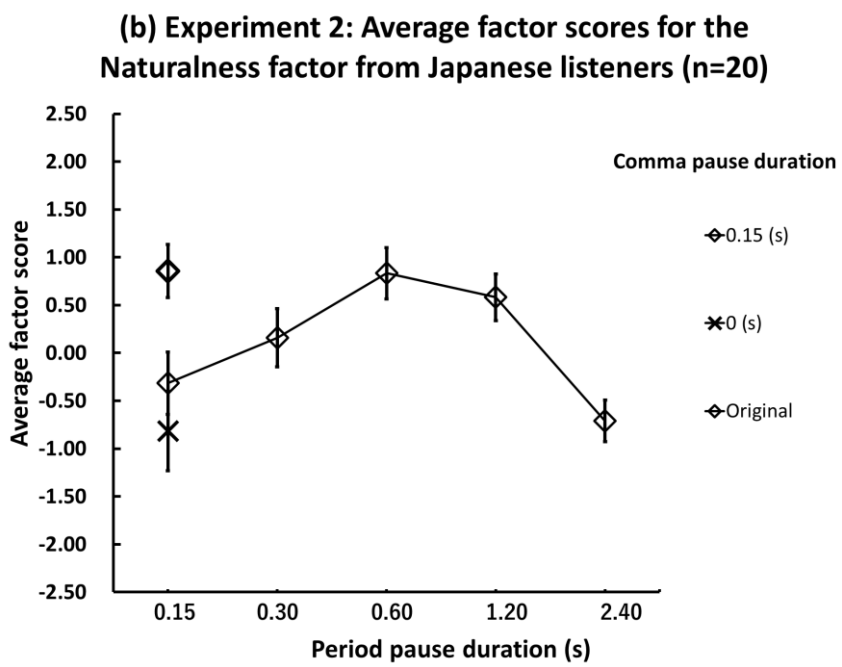
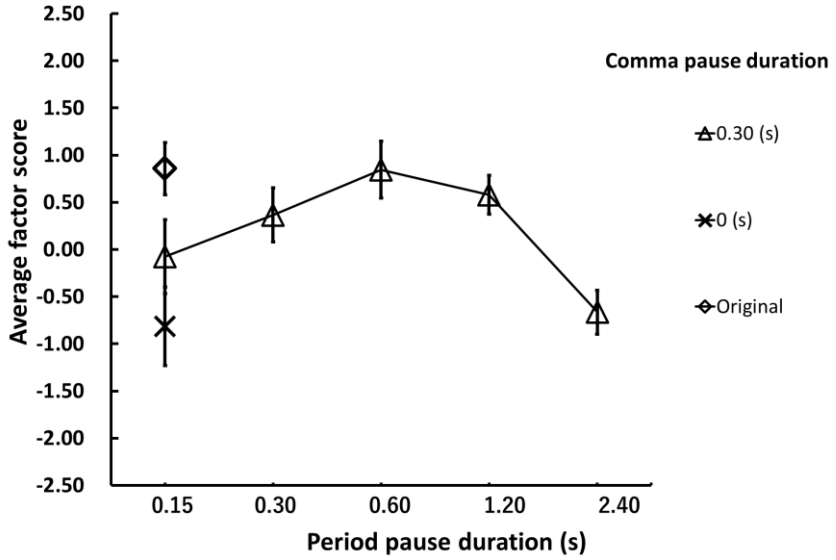


Figure 3.11 Results of Experiment 2. (a) The average factor scores for the Naturalness factor from the Chinese language group. (b) – (f) The average factor scores for the Naturalness factor for comma-pause durations of 0.15 s, 0.3 s, 0.6 s 1.2 s, and 2.4 s, respectively. The error bars show the 95%-confidence intervals.

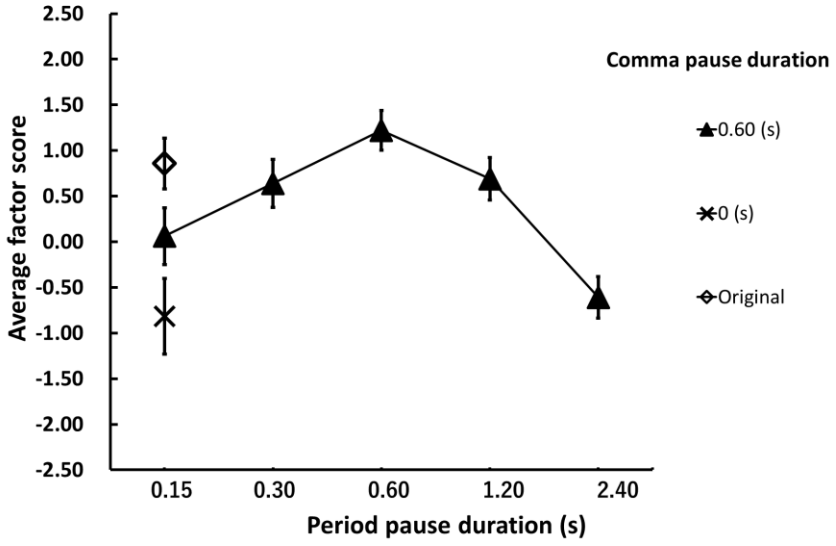

(a) Experiment 2: Average factor scores for the Naturalness factor from Japanese listeners (n=20)
Average factor score
2.50
2.00
1.50
1.00
0.50
0.00
-0.50
-1.00
-1.50
-2.00
-2.50
0.15
0.30
0.60
1.20
2.40
Period pause duration (s)
Comma pause duration
0.15 (s)
0.30 (s)
0.60 (s)
1.20 (s)
2.40 (s)
0 (s)
Original

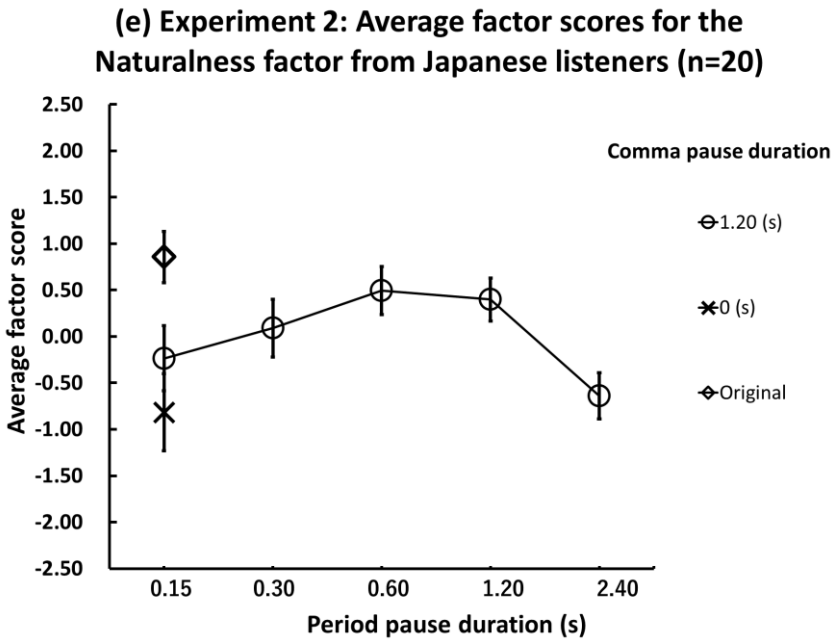

(b) Experiment 2: Average factor scores for the Naturalness factor from Japanese listeners (n=20)
Average factor score
2.50
2.00
1.50
1.00
0.50
0.00
-0.50
-1.00
-1.50
-2.00
-2.50
0.15
0.30
0.60
1.20
2.40
Period pause duration (s)
Comma pause duration
0.15 (s)
0 (s)
Original

(c) Experiment 2: Average factor scores for the Naturalness factor from Japanese listeners (n=20)



(d) Experiment 2: Average factor scores for the Naturalness factor from Japanese listeners (n=20)





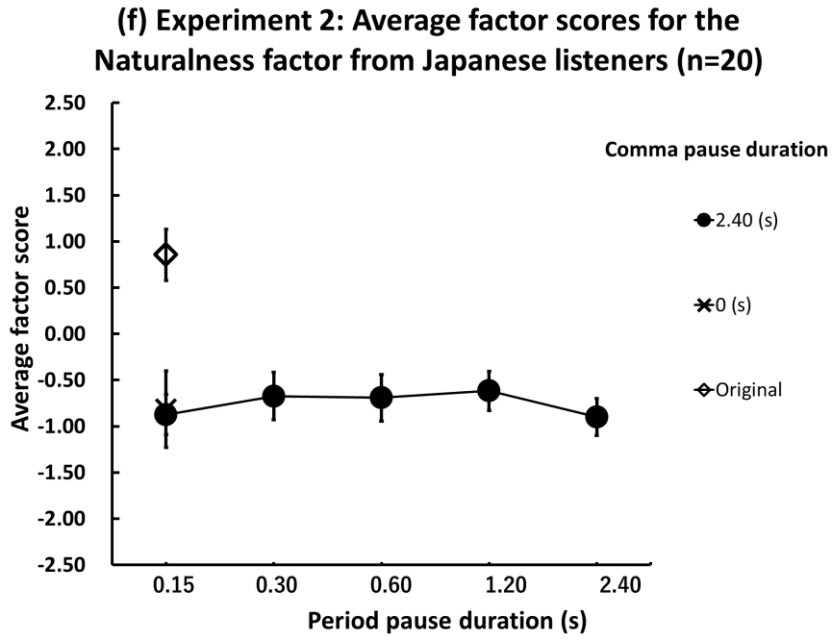


Figure 3.12 Results of Experiment 2. (a) The average factor scores for the Naturalness factor from the Japanese language group. (b) – (f) The average factor scores for the Naturalness factor for comma-pause durations of 0.15 s, 0.3 s, 0.6 s 1.2 s, and 2.4 s, respectively. The error bars show the 95%-confidence intervals.

3.4 Discussion

In this experiment, comma- and period-pause durations were varied independently. In the same way as in Experiment 1, factor analysis yielded two main factors, the Speech Rate factor and the Naturalness factor. The average factor scores for the Speech Rate factor showed that even if the speed of the utterances was not physically changed, the perceived speed of the sentences was affected by just changing the pause duration. If the pause duration lengthened, the listeners would hear slowed-down speech. The factor scores of the original speech clearly reflected the physical comma- and period-pause durations as measured in the original speech segments, showing the validity of the present analysis consisting of factor analysis over rating scale data.

Figures 3.10 - 3.12, showing factor scores for the Naturalness factor, indicate that a number of conditions of pause durations yielded speech that was as natural as the original speech. Depending on the language group, the highest scores were obtained when the comma- and period-pause duration were at a ratio of 1:2 (0.3:0.6 s; 0.6:1.2 s) or at a ratio of 1:1 (0.6:0.6 s). Thus, remarkably, even though the 1:2 ratio predominantly occurs in natural speech (O'Connell and Kowal, 1986; Yamashita and Fuyuno, 2015; Liu et al., 2016; Yamashita et al., 2019), as also is the case in our original speech segments (see **Table 1.1**), the same degree of speech naturalness could be obtained with the same comma- and period-pause duration of 0.6 s. Factor scores for this condition did not significantly differ from those for the original speech. In fact, for the English and the Japanese group, the factor scores for the 0.6-s pause duration conditions were even higher than those for the original speech, which seems to suggest common cross-cultural preference for this pause duration in English speech.

Chapter 4 – General discussion and conclusion

4.1 Summary of the results of Experiment 1 and Experiment 2

In the present thesis, two listening experiments were carried out to investigate influences of pause duration on speech impressions of English speech segments, which originally had physical pause durations typical for spoken English (**Table 2.2**). In Experiment 1, all pauses were removed, and after that all punctuated pauses (comma- and period-pauses) were remade with the same duration, between 0.075-4.8 s. In Experiment 2, comma- and period-pause durations were varied independently, in between 0.15-2.4 s. These manipulations did not change the acoustic information, but only changed the pause duration at punctuation marks. Speech segments without any pauses, as well as the original speech segments, were included as control conditions. Both native- and non-native English (Chinese and Japanese) speakers rated the segments on a broad range of items, in which the evaluation scores were first observed directly and then subjected to factor analysis.

In both experiments, three types of distribution could be observed in the evaluation data in all language groups. The three distribution types were: “curve type”, “steady decrease type”, and “no obvious change type” distribution, respectively. After factor analysis, the same two factors were observed. These factors were interpreted as representing speech naturalness and speech rate. It is to be noted that the same two factors appeared in our preliminary study with Mandarin Chinese as evaluated by native-Chinese speakers (Lin et al., 2021) as well. With regard to our research questions, the following was examined in Experiment 1: (1) how manipulated pause durations into the same durations (0.075, 0.15, 0.3, 0.6, 1.2, 2.4, and 4.8 s) at punctuation marks influenced the

subjective impression of English speech; and (2) whether there was a common pause duration for favorable impressions in English speech in different language groups. The following research questions were investigated in Experiment 2: (1) how the listener's subjective impression would change when the comma-pause and period-pause durations were varied independently in a full-factorial design (0.15, 0.3, 0.6, 1.2, and 2.4 s); and (2) whether again a similar pause duration for favorable impressions in English speech occurred for both native and non-native speakers of English (Chinese and Japanese). We observed the following. First, the perceived speech rate decreased when the physical pause duration increased. This is in line with the results of reading task experiments, in which the speech rate and the frequency and duration of pauses are interdependent (Grosjean & Collins, 1979). The results are also in line with those from a study on pause function in the production and perception of Japanese discourse (Sugito, 1990). Also in Japanese discourse, speech without pause sounded fast-paced, and changing the pause duration influenced the listeners' perception of speech rate.

Second, although the physical comma- and period-pause duration in natural speech is typically 1:2 (O'Connell & Kowal, 1986; Yamashita & Fuyuno, 2015; Liu et al., 2016; Yamashita et al., 2019), the factor scores for the Naturalness factor showed that even when the comma- and period-pause duration were equal (= 0.6 s), naturalness was very similar to – or even higher than – that of the original speech for all three language groups. In studies of time perception, durations around 0.6-0.7 s are considered as neither long nor short (Fraisse, 1964). Perceptually, the pause duration of 0.6 s therefore might be considered as natural also in English speech, regardless of the listener's language background.

4.2 About the perception of naturalness

It is remarkable that the Speech Naturalness showed a peak point in a stable manner, and how it is perceived is discussed as below.

First, according to the contents of the evaluation items “elegant”, “skillful”, “smooth”, “with appropriate rhythm”, “natural”, “experienced”, “well-practiced”, “with appropriate pause duration”, “at a suitable tempo”, “polite”, “friendly”, “fluent”, “intelligible”, and “easy to understand”, which appeared with the Naturalness factor, the factor seems related to the perception of something positive in the context of speech rhythm perception.

Second, the speech sentences that got relatively higher scores in the listening experiments had very similar mean (manipulated) pause durations as in those in the original speeches. In Experiment 1, speech sentences with a pause duration of 0.6 s got the highest factor scores in all language groups when all pause durations were manipulated to the same durations. The mean pause duration in Speech Segment 1 and Speech Segment 2 was 0.55 s (0.23), and 0.78 s (0.33), respectively. These values are close to 0.6 s. Meanwhile, in Experiment 2, (1) speech sentences with a comma-pause of 0.3 s and a period-pause of 0.6 s; (2) speech sentences with a comma-pause duration of 0.6 s and a period-pause duration of 1.2 s; and (3) speech sentences with a comma- and period-pause duration of 0.6 s, got relatively higher factor scores. These durations and the comma-period duration ratio is similar to that in the original speeches, in which comma durations varied from 0.51 s to 0.60 s, periods varied from 1.40 s to 1.43 s.

Third, the speech sentences with the relatively higher naturalness factor scores in the listening experiments had pause durations that were similar to those physically observed in speech in previous studies, typically in public speaking. For example, the

average comma and period durations in university commencement speeches in English was 0.49 s (SD = 0.26) for commas, and 1.01 s (SD = 0.40) for periods (Yamashita and Fuyuno, 2015). Furthermore, public presentations in English showed an average pause duration for commas from 0.38 to 0.45 s (SD =0.22 to 0.31), and for periods from 0.81 to 0.98 s (SD =0.31 to 0.33; Yamashita et al., 2019).

Fourth, in sentences that yielded relatively higher naturalness factor scores, the ratio between commas (0.30, and 0.60 s) and periods (0.60, and 1.20 s) was not too far from the ratio observed in previous studies [mean physical durations of commas (range: 0.38 to 0.67 s) and periods (range: 0.81 to 1.24 s); (O'Connell & Kowal, 1986; Yamashita & Fuyuno, 2015; Liu et al., 2016; Yamashita et al., 2019); That is, the ratio was close to 1:2.

4.3 Comparison with Mandarin Chinese

Lin et al (2021) conducted a similar perception experiment using Mandarin Chinese based on the method of Chapter 2 and Chapter 3 in this thesis. Both differences and similarities were observed between their study and the present study. As for the similarities, firstly, three types of distributions (i.e., a "curve type" distribution, a "steady decrease type" distribution, and a "no obvious change type" distribution) were observed. Secondly, after factor analysis was performed, the first two factors highly contributed to the squared loading, thus were taken for further consideration. These two factors could be named as the Naturalness factor and the Speech Rate factor. Thirdly, as for the Speech Rate factor, the factor scores decreased steadily as pause duration increased from 0.075 to 4.8 s in Experiment 1, and 0.15 to 2.4s in Experiment 2 for the listening experiments of both English speech and Mandarin Chinese speech (Lin et al, 2021). Moreover, as for

the Naturalness factor, the speech with all pause durations (for both commas and periods) manipulated to 0.6 s seemed to give practically an optimum condition, for both English speech and Mandarin Chinese speech (Lin et al, 2021).

However, as for the differences, the average factor score of the Naturalness factor for English original speeches (0.82 in Experiment 2) in the present study were lower than the average factor scores for Mandarin Chinese [1.27 in Experiment 1, and 1.25 in Experiment 2, respectively for Mandarin Chinese (Lin et al., 2021)]. The reasons for the difference could be as follows. First, there is a difference in the writing system. English belongs to an alphabetic writing system, which provides readers with unambiguous markers (e.g., interword spaces) to segment sentences into individual words; however, Mandarin Chinese belongs to an ideographic writing system, in which there are no unambiguous markers to segment words. Interword spaces do not exist neither in individual characters nor in words (Liu et al., 2013). Another reason for the difference in factor scores could be that the speech sources were different. Lin et al., (2021) used announcements to make the stimuli, while in the present thesis, public speeches were used. Besides, in Lin et al's study (2021), only commas and periods were included in the writing scripts, while other punctuation marks (i.e., semicolon, dash) were included in the present thesis. Last but not the least, the speaking rates of the original speech stimuli used for each language were different: In the original English speech segments, the articulation rate varied from 3.04 to 3.96 syllables per second, and the speech rate varied from 1.86 to 2.66 words per second, respectively. On the other hand, in the original Chinese speech segments, the articulation rate varied from 2.73 to 3.81 syllables per second, and the speech rate varied from 3.21 to 4.38 characters per second, respectively. Moreover, the listening experiment of Mandarin Chinese speech only employed native Chinese speakers,

however, both native and non-native (Chinese and Japanese) speakers of English were employed for the present English speech experiments. More experiments are required with data from Mandarin Chinese speeches to clarify this issue.

4.4 About the influence of language background on the results

In both of the listening experiments, native and non-native (Chinese and Japanese) speakers of English were employed to investigate whether there was a common pause duration for favorable impressions in English speech in different language groups. Overall, similarities were observed in all language groups. Two factors (i.e., Naturalness factor, Speech Rate factor) were always found in both experiments. Interestingly, speeches with 0.6 s pause duration got the highest naturalness score in all three groups. However, for the Chinese group, the factor score for the 0.3-s stimuli was not significantly lower than that for the 0.6-s stimuli in Experiment 1, whereas the difference was significant for the Japanese group. For the Japanese group, the factor score for the 0.6-s stimuli was not significantly higher than that for the 1.2-s stimuli. It is possible that the English proficiency level of the participants played a role. All the Chinese speakers got an English proficiency test score, but not all of the Japanese participants did. According to the questionnaire about the participants' English level and education, the English education system in China and Japan is different: In China, English education usually starts from primary school or even earlier, however, in Japan, English education usually starts from middle school. Moreover, the study period of English for Chinese speakers was about three years longer for Japanese speakers. This might suggest that more proficient speakers needed a relatively shorter pause duration than less proficient speakers. In further research, to investigate the influences of language backgrounds, speakers who

have different scores on the same English proficiency test could be involved.

We checked the results with this hypothesis in Experiment 2. If so, the Chinese participants should prefer a relatively shorter pause duration than Japanese participants did. Surprisingly, for the Chinese language group, the speech sentences with a comma-pause duration of 0.6 s and a period-pause duration of 1.2 s (0.79) got the second highest naturalness factor score compared to the original speech (0.81). For the Japanese language group, the factor scores for speech sentences with a comma- and period-pause duration of 0.6 s was the highest (1.2), again, even exceeding the factor score obtained for the original speech (0.85). In Experiment 2, all the Chinese participants got English test proficiency scores, but not all the Japanese participants did. Thus, the hypothesis was not confirmed in Experiment 2. Besides, our study with Mandarin Chinese showed that speech segments with a comma-pause duration of 0.6 s, along with a period-pause duration of 0.6 s or 1.2 s, received the highest scores for the Naturalness factor, and these were not significantly different from the factor scores for the original speech (Lin et al, 2021). Thus, in the listening experiment in which pause duration was varied independently, it seemed that Chinese participants preferred speech with the comma-pause duration of 0.6 s, and a period-pause duration of 1.2 s, regardless of the target language.

4.5 Feasibility to implement factor analysis

In a study of description of personality, Goldberg (1990) built the Big-Five factor structure, in which the personality could be clarified as extraversion, neuroticism, openness to experience, conscientiousness, and agreeableness. Uchida did several studies between the acoustic information and personality based on the method of Goldberg

(1990). For example, the effects of the speech rate conversation on the impressions of pitch and the images of speakers' personality were investigated by Uchida (2000). Effects of the speech rate on speakers' personality-trait impressions, and effects of fundamental frequency and speech rate conversion upon impressions of talkers' personality were also investigated by Uchida (2002, 2003). Moreover, effects of speech rate and pause duration on impression of speakers' personality and the naturalistic qualities of speech were investigated by Uchida (2005). These studies formed a basis to utilize a similar method to investigate how pause duration influences the subjective impressions of English speech.

In order to check whether factor analysis could be implemented on the large-scale rating data of the present thesis, we first observed the distribution of the original rating scores on each evaluation item. As a result, overall, three types of distributions were found in both of the experiments. A "curve type" distribution, which had relatively low evaluation scores with short or long pause durations, but relatively high evaluation scores for intermediate pause durations; an approximately "steady decrease type", in which the evaluation scores decreased steadily as the pause duration became longer; and a "no obvious change type", in which the evaluation scores did not change obviously as the pause duration changed. Interestingly, the three types of distribution were also found in different language groups of listeners. Moreover, in the two listening experiments of Mandarin Chinese evaluated by Chinese speakers, the three types of distributions were also observed (Lin et al., 2021).

Comparing the analysis method and results with Uchida's (2005), similarities and differences were observed. As for the differences, Uchida (2005) only performed principal component analysis (PCA) on the rating data, as a result, 4 principal components were extracted. According to the percentage of variance from the eigenvalues, the highest

two principal components were taken for further consideration. However, in the present thesis, factor analysis was performed, and four factors were extracted with varimax rotation after principal component analysis (PCA). The first and the second factor were taken into further consideration, because their cumulative percentage of variance in the rotation sums of squares was about 60%. As for the similarities, the first principal component or factor could be named as Speech Naturalness, the second principal component or factor could be named as Speech Rate, respectively. This suggests the feasibility to implement factor analysis on the present large-scale rating data, moreover, regardless of the listeners' language background and the target languages.

4.6 Further research and limitations of the present study

We anticipate that the perceived naturalness at an equal comma- and period-duration of 0.6 s is of use in training L2-speakers of English, for example those whose native tongue is not a stress language, because they can simply be instructed to use the same pause duration when delivering speeches in English; pausing is easier to acquire and to control than pronunciation (Matzinger et al., 2020). Furthermore, the present results may assist developments in artificial speech technology, regarding both speech generation and recognition. Further research, however, is necessary in order to clarify whether the 0.6-s pause duration is natural for other languages as well.

Our study is also limited in that pause durations in other syllable-based languages or a mora-based language (Japanese) need to be investigated as well. Finally, it is still unclear whether the natural pause duration depends on the difficulty level of the English content, or on whether the speaker is a native English speaker or not.

4.7 Conclusion of the present thesis

Overall, this thesis research was conducted to investigate how manipulated pause duration at punctuation marks influences the subjective impression of English speech. It is expected that the findings of this thesis research will be useful for research related to speech perception, and English education.

The present research showed that:

(a) Three types of distributions (i.e., a "curve type" distribution, a "steady decrease type" distribution, and a "no obvious change type" distribution) were observed in the evaluation data in speech sentences according to the trends in the evaluation scores for all items (see **Table 2.1**) in all language groups (English, Chinese, Japanese). Thus, it provided the possibility to subject the rating data to a factor analysis.

(b) The same two factors were observed in all language groups, namely, the Naturalness factor and the Speech Rate factor. The factor scores of the Speech Rate factor decreased as manipulated pause duration increased for all language groups, even though the same acoustic information in the utterance segment was given to the listeners. In other words, pause duration influenced the speech rate. Next, speeches with a pause duration of 0.6 s got the highest Naturalness factor scores in all language groups when all pause durations were manipulated to the same durations from 0.075 to 4.8 s. For the native-English group, the factor score was significantly higher than that for any of the other stimuli. For the Chinese group, the factor score for the 0.6-s stimuli was not significantly higher than that for the 0.3-s stimuli. For the Japanese group, the factor score for the 0.6-s stimuli was not significantly higher than that for the 1.2-s stimuli. When pause durations were manipulated separately for commas and periods, there were no significant differences between the factor scores for the original

speech and (1) speech sentences with a comma-pause of 0.3 s and a period-pause of 0.6 s; (2) speech sentences with a comma-pause duration of 0.6 s and a period-pause duration of 1.2 s; and (3) speech sentences with a comma- and period-pause duration of 0.6 s. Thus, pause duration influenced the naturalness of speech, and a pause duration of 0.6 s seemed to be a commonly favorable pause duration in different language backgrounds.

Together, based on the findings with manipulated pause durations in English speech, this thesis suggests that pause duration influences human subjective impressions of speech mainly in two dimensions: the naturalness of speech and the speech rate. The pause duration of 0.6 s occurs as a common pause duration that promotes naturalness in English speech, in different language groups.

References

Bai, X., Yan, G., Liversedge, S. P., Zang, C., & Rayner, K. (2008). Reading spaced and unspaced Chinese text: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance,* 34(5), 1277–1287. https://doi.org/10.1037/0096-1523.34.5.1277

Barry, K. (2017). *Pause: Standout Whenever You Speak.* Pennsauken Township: BookBaby. (Self-published)

Beijin Commercial Press, & Shogakukan (2016). *Chinese-Japanese Dictionary,* 3rd Edn. Tokyo: Shogakukan.

Black, J. W., Tosi, O., Singh, S., & Takefuta, Y. (1966). A study of pauses in oral reading of one's native language and in English. *Language and Speech,* 9, 237-241. https://doi.org/10.1177/002383096600900405

Boersma, P., & Weenink, D. (2015). Praat: doing phonetics by computer [Computer program]. Version 6.0.21, retrieved January 2015 from https://www.praat.org/.

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics,* 27, 207-229. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.472.7571&rep=rep1&type=pdf

Croes, J. K., & Dexter, K. (2016). 「『*The Art of*、 *Japanese Punctuation*～。』！？. (net-published) https://www.tofugu.com/japanese/japanese-punctuation/

Cruttenden, A. (1986). *Intonation.* Cambridge University Press.

Crystal, D. (2008). Two thousand million? Updates on the statistics of English. *English Today*, 24(1), 3-6. https://doi.org/10.1017/S0266078408000023

Dawadee, P., Prabhu, S., & Bhattarai, B. (2017). Wordlist in nepali to assess articulatory errors. *Nepalese Journal of ENT Head and Neck Surgery,* 7(2),11-6. https://www.njehns.org.np/index.php/njehns/article/view/169

de Johnson, T. H., O'Connell, D. C., & Sabin, E. J. (1979). Temporal analysis of English and Spanish narratives. Bull. Psychon. Soc. 13, 347–350. doi: 10.3758/ BF03336891

Dorman, M. F., Raphael, L. J., & Liberman, A. M. (1979). Some experiments on the sound of silence in phonetic perception. *The Journal of the Acoustical Society of America*, 65(6), 1518-1532. doi: 10.1121/1.382916. PMID: 489822.

Faculty of Liberal Arts, University of Tokyo English Subcommittee (1998). *The Universe of English II.* Tokyo: University of Tokyo Press.

Faculty of Liberal Arts, University of Tokyo English Subcommittee (2000). *The Expanding Universe of English II.* Tokyo: University of Tokyo Press.

Field, A. (2009). *Non-parametric Tests.* In Discovering Statistics Using SPSS, 3rd edition, Sage Publication: London, United Kingdom.

Fraisse, P. (1964). *The Psychology of Time.* (Jennifer Leith, M. A, Trans.). Eyre & Spottiswoode.

Fraisse, P. (1982). Rhythm and Tempo. In D. Deutsch (Ed.), *The Psychology of Music* (pp. 149–180). Academic Press.

Fuyuno, M., Yamashita, Y., & Nakajima, Y. (2016). Multimodal corpora of English public speaking by Asian learners: Analyses on speech rate, pause and head gesture. In F. A. Almeida, I. O. Barrera, E. Q. Toledo, & M. S. Cuervo (Eds.), *Input a Word, Analyse the World: Selected Approaches to Corpus Linguistics* (pp. 461-476). Cambridge Scholars Publishing.

Fuyuno, M., Yamashita, Y., Saitoh, T., & Nakajima, Y. (2017). Semantic structure, speech units and facial movements: Multimodal corpus analysis of English public speaking. In A. M. Ortiz, & C. Pérez-Hernández. *EPiC Series in Language and Linguistics,* 1, 447-461. https://doi.org/10.29007/tmcf

General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China & Standardization Administration. (2011). General rules for punctuation (GB/T 15834-2011). Beijing: Standard Press of China.

Gilakjani, P. A. (2012). A study of factors affecting EFL Learners' English pronunciation learning and the strategies for instruction. *International Journal of Humanities and Social Science,* 2(3), 119-128.

Goldberg, L. R. (1990). An alternative "Description of Personality": The Big-Five factor structure. *Journal of Personality and Social Psychology,* 59, 1216-1229. https://doi.org/10.1037/0022-3514.59.6.1216

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in Spontaneous Spoken Discourse*. London: Academic Press.

Goldman-Eisler, F. (1972). Pauses, clauses, sentences. *Language and Speech,* 15, 103-113. https://doi.org/10.1177/002383097201500201

Goto, M., Itou, K., & Hayamizu, S. (1999). A real-time filled pause detection system for spontaneous speech recognition. In F. McInnes, D. Attwater, M. Edgington, M.S. Schmidt, & M. Jack. *Proceedings of the 6th European Conference on Speech Communication and Technology,* Budapest, Hungary (pp. 227-230). Eurospeech 99. https://staff.aist.go.jp/m.goto/PAPER/EUROSPEECH99.pdf

Grosjean, F., & Collins, M. (1979). Breathing, pausing and reading. Phonetica, 36(2), 98-114. https://doi.org/10.1159/000259950

Grosjean, F. H., Grosjean, L., & Lane, H. (1979). The patterns of silence: Performance structures in sentence production. *Cognitive Psychology*, 11(1), 58-81. https://doi.org/10.1016/0010-0285(79)90004-5

Handel, S. (1989). *Listening: An Introduction to the Perception of Auditory Events.* The MIT Press.

Heldner, M., & Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, 38(4), 555–568. https://doi.org/10.1016/j.wocn.2010.08.002

Hieke, A. E., Kowal, S., & O'Connell, D. C. (1983). The trouble with "articulatory" pauses. *Language and Speech*, 26(3), 203–214. https://doi.org/10.1177/002383098302600302

Horii, Y. (1983). An automatic analysis method of utterance and pause length and frequencies. *Behavior Research Methods & Instrumentation*, 15(4), 449-452. https://link.springer.com/content/pdf/10.3758/BF03203682.pdf

Jacewicz, E., Fox, R., O'Neill, C., & Salmons, J. (2009). Articulation rate across dialect, age, and gender. *Language Variation and Change,* 21(2), 233–256. https://doi.org/10.1017/S0954394509990093

Jenkins, J., Cogo, A., & Dewey, M. (2011). Review of developments in research into English as a lingua franca. *Language Teaching,* 44(3), 281-315. https://doi.org/10.1017/S0261444811000115

Kang, Y. (2014). Voice Onset Time merger and development of tonal contrast in Seoul Korean stops: A corpus study. *Journal of Phonetics,* 45,76-90. https://doi.org/10.1016/j.wocn.2014.03.005

Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language,* 60, 286-319. DOI: https://doi.org/10.2307/413642

Koch, A. (2010). *Speaking With a Purpose* (8th ed.). Boston, MA: Allyn & Bacon

Kohno, M. (1998). Mora, Syllable, and Rhythm - A Psycholinguistic Study. *Journal of the Phonetic Society of Japan,* 2(1), 16-24. https://www.jstage.jst.go.jp/article/onseikenkyu/2/1/2_KJ00007630824/_pdf

Kondo, I., & Takano, F. (Eds.). (2002). *Shogakukan Progressive Japanese-English Dictionary,* 4th Edn. Tokyo: Shogakukan.

Konishi, T., & Minamide, K. (Eds.). (2001). *Taishukan's Unabridged Genius English-Japanese Dictionary.* Tokyo: Taishukan Publishing Company.

Kowal, S., Wiese, S., & O'Connell, D. (1983). The use of time in storytelling. *Language and Speech,* 26(4), 377-392. https://doi.org/10.1177/002383098302600405

Kubozono, H. (2015). I Introduction to Japanese phonetics and phonology. In H. Kubozono (Ed.), *Handbook of Japanese Phonetics and Phonology* (pp. 1-40). Berlin, München, Boston: De Gruyter Mouton. https://doi.org/10.1515/9781614511984.1

Kuhnert, B., & Antolík, T. (2018). Patterns of articulation rate in English / French tandem interactions. The pronunciation of English by speakers of other languages. In J. Volín & R. Skarnitzl (Eds.), *The Pronunciation of English by Speakers of Other Languages* (pp. 210-226). Cambridge Scholar Publishing.

Ladefoged, P., & Johnson, K. (2006). *A Course in Phonetics* (6th ed.). Cengage Learning. https://vulms.vu.edu.pk/Courses/ENG507/Downloads/A-Course-in-Phonetics.pdf

Ladefoged, P., & Johnson, K. (2015). *A Course in Phonetics* (7th ed.). Cengage Learning.

Li, F. (2013). The effect of speakers' sex on voice onset time in Mandarin stops. *The Journal of the Acoustical Society of America,* 133(2), 142-147. https://doi.org/10.1121/1.4778281

Liberman, A. M. (1996). *Speech: A special code.* The MIT Press.

Lin, M., Nakajima, Y., Liu, S., Ueda, K., & Remijn, G.B. (2021). The influence of comma- and period-pause duration on the listener's impression of speeches made in mandarin Chinese. *The Asian Conference on Language 2021: Official Conference Proceedings,* Tokyo, Japan (pp. 209-216).

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: acoustical measurements. *Word,* 20(3), 384-422. https://doi.org/10.1080/00437956.1964.11659830

Liu, P., Li, W., Lin, N., & Li, X. (2013). Do Chinese readers follow the national standard rules for word segmentation during reading? *PLoS One,* 8(2): e55440. https://doi.org/10.1371/journal.pone.0055440

Liu, S., Nakajima, Y., & Yamashita, Y. (2016). Analysis of temporal structure of English speech in public speaking presented by Japanese EFL learners. *Proceedings of the Auditory Research Meeting 46(9),* Fukuoka, Japan (pp. 565-570).

Lucas, S. E. (2015). *The Art of Public Speaking* (12th ed.). McGraw-Hill Education.

Malt, B. C., & Seamon, J. G. (1978). Peripheral and cognitive components of eye guidance in filled-space reading. *Perception & Psychophysics, 23*(5), 399–402. https://doi.org/10.3758/BF03204142

Matzinger, T., Ritt, N., & Fitch, W.T. (2020). Non-native speaker pause patterns closely correspond to those of native speakers at different speech rates. *PLoS One,* 15(4), e0230710. https://doi.org/10.1371/journal.pone.0230710

McCawley, J. D. (1968). *The Phonological Component of a Grammar of Japanese*. The Hague: Mouton.

McCawley, J. D. (1978). What is a tone language. In V. Fromkin (Eds.), *Tone: A linguistic Survey* (pp. 113-131). New York, NY: Academic Press.

NTT-AT. (2002). Multi-lingual speech database 2002.

O'Connell, D.C., & Kowal, S. (1983). Pausology. In W. A. Sedelow & S. Y. Sedelow (Eds.), *Computers in language research 2* (pp. 221-301). Mouton Publishers.

O'Connell, D.C., & Kowal, S. (1984). Comparisons of native and foreign language poetry readings: Fluency, expressiveness, and their evaluation. *Psychological Research,* 46(3), 301-313. https://link.springer.com/content/pdf/10.1007/BF00308890.pdf

O'Connell, D.C., & Kowal, S. (1986). Use of punctuation for pausing: Oral readings by German radio homilists. *Psychological Research,* 48, 93-98. https://doi.org/10.1007/BF00309205

Oehmen R., Kirsner K., Fay N. (2010). Reliability of the manual segmentation of pauses in natural speech. In H. Loftsson, F. Rögnvaldsson, & S. Helgadóttir (Eds.), *Advances in Natural Language Processing. NLP 2010. Lecture Notes in Computer Science,* 6233. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-14770-8_30

O'Hair, D., Stewart, R., & Rubenstein, H. (2001). *A Speaker's Guidebook: Text and Reference*. Boston, MA: Bedford/St. Martin's.

Oliveira, M. (2002). The role of pause occurrence and pause duration in the signaling of narrative structure. In E. Ranchhod, & N. J. Mamede. *Proceedings of the Third International Conference on Advances in Natural Language Processing (PorTAL '02),* Faro, Portugal (pp. 43–52). Springer. https://link.springer.com/content/pdf/10.1007%2F3-540-45433-0_7.pdf

Pett, M. A., Lackey, N. R., & Sullivan, J. J. (2003). *Making Sense of Factor Analysis.* Thousand Oaks, Calif: Sage Pub. https://www.doi.org/10.4135/9781412984898

Pike, K. (1945). *The Intonation of American English.* Ann Arbor: University of Michigan Press.

Rosen, K., Murdoch, B., Folke, J., Vogel, A., Cahill, L., Delatycki, M., & Corben, L. (2010). Automatic method of pause measurement for normal and dysarthric speech. *Clinical Linguistics & Phonetics,* 24(2), 141-154. https://doi.org/10.3109/02699200903440983

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematic for the organization of turn-taking for conversation. *Language,* 50 (4), 696-735

Straus, J., Kaufman, L., & Stern, T. (2014). *The Blue Book of Grammar and Punctuation* (11th ed.). Wiley.

Stuckenberg, A., & O'Connell, D.C. (1988). The long and short of it: reports of pause occurrence and duration in speech. *Journal of Psycholinguistic Research,* 17(1), 19-28. https://doi.org/10.1007/BF01067179

Suen, C. Y. & Beddoes, M.P. (1974). The silent interval of stop consonants, *Language & Speech,* 17, 126-134.

Sugito, M. (1990). On the role of pauses in production and perception of discourse. *Conference Proceedings, First International Conference on Spoken Language Processing (ICSLP 1990),* Kobe, Japan, (pp. 513-516). https://www.isca-speech.org/archive/pdfs/icslp_1990/sugito90_icslp.pdf

Takada, M. (2008). Geographical pattern of VOT in Japanese initial voiced stops. *Studies in the Japanese Language,* 4(4), 48-62. https://www.jstage.jst.go.jp/article/nihongonokenkyu/4/4/4_KJ00005586585/_pdf

Taneichi, A. (2014). Reconsideration of silence in conversation: Toward a typology of silence as pragmatic acts. *Intercultural Communication Review.* 12, 145-156. http://doi.org/10.14992/00011117

Todd, N. (1985). A model of expressive timing in tonal music. *Music Perception,* 3(1), 33-57. https://doi.org/10.2307/40285321

Tosi, O.I. (1965). *A method for acoustic segmentation of continuous sound into pauses and signals and measurement of segment durations* [Unpublished PhD thesis]. Ohio State University.

Uchida, T. (2000). Effects of the speech rate conversion on the impressions of pitch and the images of speakers' personality. *Journal of the Acoustical Society of Japan*, 56, 396-405.)

Uchida, T. (2002). Effects of the speech rate on speakers' personality-trait impressions. *Japanese Journal of Psychology*, 73, 131-139. https://doi.org/10.4992/jjpsy.73.131

Uchida, T. (2003). The effects of fundamental frequency and speech-rate conversion upon impressions of talkers' personality. In C. Don (Ed.), *Acoustics on the Move: Proceedings of the Eighth Western Pacific Acoustics Conference: WESPAC VIII*, Castlemaine: Australian Acoustical Society, Speech-TB23.

Uchida, T. (2005). Impression of speaker's personality and the naturalistic qualities of speech: speech rate and pause duration. *The Japanese Journal of Educational Psychology*, 53(1), 1-13. https://doi.org/10.5926/jjep1953.53.1_1

University of International Business and Economics, Shogakukan, & Beijin Commercial Press (2015). *Japanese-Chinese Dictionary,* 3rd Edn. Tokyo: Shogakukan.

Wrench, S. J., Goding, A., Johnson, I. D., & Attias, A. B. (2016). *Stand Up, Speak Out: The Practice and Ethics of Public Speaking.* FlatWorld.

Yamashita, Y., Fuyuno, M. & Nakajima, Y. (2014). Influence of speech rate and pauses on the efficiency of English public speaking of Japanese EFL learners. *Proceedings of the Auditory Research Meeting 44(9),* Fukuoka, Japan (pp. 561-564).

Yamashita, Y., & Fuyuno, M. (2015). An analysis of speech pauses, head movements, and lexical choices in graduation speeches at the universities in the United States. *ATEM Journal: Teaching English Through Movies*, 20, 33-44. https://doi.org/10.24499/atem.20.0_33

Yamashita, Y., Fuyuno, M. & Nakajima, Y., Antonijevic-Elliott, S., Elliott, M.A., & Arndt, S. (2019). A quantitative analysis to characterize verbal and non-verbal behavior on public speaking performance [Paper presentation]. In N. du Bois, S. Arndt, E. V. Özsoy, S. Bayraktar, E. Gülbetekin, & M.A. Elliott (Eds). *Conference Proceedings, 35th Annual Meeting of the International Society for Psychophysics,* Fechner Day 2019, 35th Annual Meeting of the International Society for Psychophysics, Antalya, Turkey (pp. 13).

Appendix A. Informed consent and instructions of Experiment 1

KYUSHU UNIVERSITY

Human Science, Faculty of Design
4-9-1, Shiobaru, Minami-ku, Fukuoka
Japan
Prof Dr. Yoshitaka Nakajima

Consent Form

I, ...……....., give consent to my participation in the research
Name (please print)

project entitled **"Human perception and cognition: Analysis of speech from the viewpoint of perceptual psychology"**

In giving my consent, I acknowledge that:

1. The procedures required for the project and the time involved have been explained to me, and any questions I have about the project have been answered to my satisfaction.

2. I have read the Participant Information Statement and have been given the opportunity to discuss the information and involvement in the project with the researcher/s.

3. I understand that participation is voluntary and that I can withdraw from the study at any time, without penalty or prejudice.

4. I understand that my involvement is strictly confidential and no information about me and my data will be used in any way that reveals my identity, except when agreed otherwise.

…… Date …….
Full name (please print)

If you require further information, please contact :

Yoshitaka Nakajima
Telephone: +81 92 553 4564 / Fax: +81 92 553 4520
Address: 4-9-1, Shiobaru, Minami-ku, Fukuoka, 815-8540, Japan
Email: nakajima@design.kyushu-u.ac.jp

Instructions

Thank you for joining the experiment today.
From now on, English speeches will be presented.

How to operate the computer program.

1. After entering the lab, and the screen will be shown as below:

2. Press the "PLAY" button and the speech will be presented.
3. Please evaluate the evaluation items according to the answer sheet prepared on the table, after hearing the speech. For the evaluation method, please refer to the following "solution paper evaluation method".
4. Please press the "NEXT" button to enter the next stimulus screen when you finish evaluate the speech.
5. Please press the "PLAY" button to display the next stimulus.
6. It means that the session is finished, when the message "Thanks for your cooperation. Please take a rest!" presented on the screen. You can take a rest if you want or go on to the next session.

Evaluation method of the answer sheet.

1. There will be 12 evaluation items for one stimulus. Please judge how it is related to the speech. An example of evaluation item is shown as below.

Stimulus 1:

	not									very much
smooth	1	2	3	4	5	6	7	8	9	10

2. There will be 10 steps in each evaluation items. “1” presents for “not”, and “10” presents for “very much”.

3. When listen to the speeches, please imagine that you are an inspector of English public speaking. After listening to the speech, please judge the content of each evaluation project based on performance (speech style) of the speech.

4. Please turn off the answer sheet after finishing judging one stimulus. Please do not refer to previous evaluation scores.

5. Please circle the score on the answer sheet. Do not circle the space between the numbers, please select a number.

6. Before the experiment, please have a look at the 12 evaluation items.

Attentions

·After pressing the play button, Please do not touch the mouse until a series of speeches stop.

· Please don't look at the watch or any other things that can show the time.

·Please don't be hesitate to tell me if you do not feel very well or make a operation mistake. The experiment will not give you any harm, even if you do not finish the whole experiment.

· Please do not use your hands or any other part of your body to take a rhythm.

· There will be a practice session, session 1 and session 2 in this experiment. You can take a break between each section.

• The next stimuli will not be presented if you just press the "NEXT" button. Please make sure to press the “PLAY” button in order to present the next stimuli.

Thanks for your cooperation.

实验指示

感谢您百忙之中抽出时间参加今天的实验。
从现在开始，将播放英语演讲包含演说，口头发表，席间致辞等。

如何操作计算机程序。

1. 进入实验室后，计算机屏幕画面如下图所示:

2. 按“PLAY”键，播放音频。

3.音频播放完后，请根据桌上准备的答题纸上的评价项目对音频进行评分。评分方法请参考以下“答题纸评分方法”。

4. 当您完成评分时，请按下“NEXT”键进入下一个音频画面。

5. 请按“播放”键，以播放下一个音频。

1. 当屏幕提示“谢谢您的合作”时，表示此环节结束。你可以适当休息一下，或者继续下一环节。

答题纸评分方法。

1. 一个音频对应有 12 个评价项目。请对每项评价项目与所播放音频的关系进行评分。评价评价项目的示例如下所示。

音频 1:

	非常不符合									非常符合
流畅的	1	2	3	4	5	6	7	8	9	10

2.每个评价项目将有 10 个阶段。“1”表示“非常不符合”,“10”表示“非常符合”。

3.听演讲音频时候，请想象你是一名英语演讲的指导老师。听完演讲音频后，请根据每项评价项目的内容对演讲的整体表现进行评分。

4. 完成一个音频的评分后，请反扣答题纸。评价当前音频时请不要参考前一个音频的评分内容。

5. 请圈答题纸上的显示的分数。不要圈数字之间的空格，请选择一个数字画圈。

6. 在实验之前，请浏览一下 12 个评价项目。

注意事项

•按下播放按钮后，请不要触摸鼠标，直到音频播放结束。

•请不要看手表或任何其他能显示时间的物品。

•如果你觉得不舒服或操作失误，不要有任何犹豫请告知主试。即使你没有完成整个实验，这个实验也不会给你带来任何伤害。

•请不要用你的手或身体的任何其他部位来敲击节奏。

•本实验将有一个练习环节，第 1 环节和第 2 环节。你可以在每个环节之间休息一下。

•如果您只按下“NEXT”按钮，音频不会自动播放。请务必按下“PLAY”按钮，以播放音频。

谢谢您的合作。

教示

本日は実験に参加していただき、ありがとうございます。
これから、英語スピーチ(演説、口頭発表、テーブルスピーチ等を含む)を呈示します。

パソコンを操作する手順

1. 実験室に入ると、パソコンのスクリーン画面で次のように表示されます。

2. 「PLAY」ボタンを押すと、音声が呈示されます。
3. 音声を聞いた後、机の上に用意してある回答用紙の評定項目に即して評定してください。評定する方法については以下の「回答用紙の評定方法」を参照してください。
4. 評定した後、「NEXT」ボタンを押すと、次の刺激の画面に入ります。
5. 「PLAY」ボタンを押すと、次の音声が呈示されます。
6. 画面に「お疲れ様でした。」というメッセージが表示されたら、セッションが終わるという意味です。休憩を取っても、次のセッションに進んでも結構です。

回答用紙を用いた評定方法

1. 一つの刺激に対して、12 項目の評定を行います。それぞれの項目がどのように音声に関わるかを判定してください。評定項目の例を以下に示します。

刺激 1：

全く当てはまらない　　非常によく当てはまる

滑らかな：　　1　2　3　4　5　6　7　8　9　10

2. 各評定項目の内容について“全く当てはまらない”(1)から“ 非常によく当てはまる”(10)まで、10 段階 （1－10）を設けています。
3. スピーチを聞く際には、自分が英語演説の指導員であると想像してください。スピーチを聞き終わってから，各評定項目の内容に即して話しかたの評定を行ってください。
4. 一つの刺激を評定し終わったら、回答用紙を裏返してください。前の評定点数を参照しないようにしてください。
5. 評定用紙に書かれた点数に○を付けてください。数字と数字の間には○を付けずに、必ず、いずれかの数字に○を付けてください。
6. 実験が始まる前に、12 の評価項目に目を通しておいてください。

注意事項

- 「PLAY」ボタンを押した後は、一連のスピーチの再生が終わるまでマウスに触らないでください。
- 時計など、時間のわかるものは見ないでください。
- 気分が優れなかったり、操作を間違えたりした場合は、遠慮なく実験者に申し出てください。実験を中断しても、実験自体に不都合が生じることはありません。
- 手足など、体の一部でリズムを取らないようにしてください。
- 本実験は、練習セッション、第 1 セッション、第 2 セッションの 3 つで構成されています。各セッションの間に休憩をとることができます。
- 「NEXT」ボタンを押しただけでは、次の音声は自動的には流れません。次の音声を流すためには、必ず「PLAY」ボタンを押してください。

それでは、よろしくお願いいたします。

アンケート

名前：

性別：　　男性　　女性

年齢：

大学：　　専攻：　　学年：

いつから英語を勉強し始めましたか？

英語を何年ぐらい勉強していますか？

英語テストの成績ありますか？点数はいくらですか（およその値で結構です）？
（TOEIC，IELTS，TOEFL，GMAT など）

大学などで、英語の授業を受けていますか？もし受けている場合、どんな授業ですか？

普段の生活において、英語を使う機会ありますか？どんな場合に英語を話しますか？

自分の英語について、どんなことが気になっていますか？改善したいことはありますか？（例えば：話す力、書く能力、読解力、文法、発音など）

パブリック・スピーキングに関する訓練を受けたことありますか？もし受けた場合、どんな訓練でしたか？

本日の評価用紙の評価項目について、感想などはありますか？

Stimulus

	not									very much
rushed	1	2	3	4	5	6	7	8	9	10
intelligible	1	2	3	4	5	6	7	8	9	10
high-pitched	1	2	3	4	5	6	7	8	9	10
skillful	1	2	3	4	5	6	7	8	9	10
polite	1	2	3	4	5	6	7	8	9	10
dynamic	1	2	3	4	5	6	7	8	9	10
clear-cut	1	2	3	4	5	6	7	8	9	10
elegant	1	2	3	4	5	6	7	8	9	10
smooth	1	2	3	4	5	6	7	8	9	10
nervous	1	2	3	4	5	6	7	8	9	10
speedy	1	2	3	4	5	6	7	8	9	10
at a suitable tempo	1	2	3	4	5	6	7	8	9	10
experienced	1	2	3	4	5	6	7	8	9	10
shrill	1	2	3	4	5	6	7	8	9	10
fluent	1	2	3	4	5	6	7	8	9	10
with appropriate pause duration	1	2	3	4	5	6	7	8	9	10
well-practiced	1	2	3	4	5	6	7	8	9	10
fast	1	2	3	4	5	6	7	8	9	10
easy to understand	1	2	3	4	5	6	7	8	9	10
rough-timbred	1	2	3	4	5	6	7	8	9	10
friendly	1	2	3	4	5	6	7	8	9	10
natural	1	2	3	4	5	6	7	8	9	10
with appropriate rhythm	1	2	3	4	5	6	7	8	9	10

刺激

	非常不同意									非常同意
明白易懂的	1	2	3	4	5	6	7	8	9	10
声调高的	1	2	3	4	5	6	7	8	9	10
擅长的	1	2	3	4	5	6	7	8	9	10
礼貌的	1	2	3	4	5	6	7	8	9	10
动感的	1	2	3	4	5	6	7	8	9	10
口齿清脆的	1	2	3	4	5	6	7	8	9	10
优雅的	1	2	3	4	5	6	7	8	9	10
流畅的	1	2	3	4	5	6	7	8	9	10
紧张的	1	2	3	4	5	6	7	8	9	10
有速度感的	1	2	3	4	5	6	7	8	9	10
节拍适当的	1	2	3	4	5	6	7	8	9	10
经验丰富的	1	2	3	4	5	6	7	8	9	10
尖锐的	1	2	3	4	5	6	7	8	9	10
流利的	1	2	3	4	5	6	7	8	9	10
停顿时长适当的	1	2	3	4	5	6	7	8	9	10
充分练习的	1	2	3	4	5	6	7	8	9	10
语速快的	1	2	3	4	5	6	7	8	9	10
容易理解的	1	2	3	4	5	6	7	8	9	10
音质粗糙的	1	2	3	4	5	6	7	8	9	10
友好的	1	2	3	4	5	6	7	8	9	10
自然的	1	2	3	4	5	6	7	8	9	10
节奏适当的	1	2	3	4	5	6	7	8	9	10
匆忙的	1	2	3	4	5	6	7	8	9	10

刺激

	全くあてはまらない									非常によくあてはまる
自然な	1	2	3	4	5	6	7	8	9	10
リズムが適切である	1	2	3	4	5	6	7	8	9	10
せきたてるような	1	2	3	4	5	6	7	8	9	10
聴き取りやすい	1	2	3	4	5	6	7	8	9	10
高音である	1	2	3	4	5	6	7	8	9	10
上手な	1	2	3	4	5	6	7	8	9	10
丁寧な	1	2	3	4	5	6	7	8	9	10
活力のある	1	2	3	4	5	6	7	8	9	10
歯切れの良い	1	2	3	4	5	6	7	8	9	10
優雅な	1	2	3	4	5	6	7	8	9	10
滑らかな	1	2	3	4	5	6	7	8	9	10
緊張感がある	1	2	3	4	5	6	7	8	9	10
スピード感がある	1	2	3	4	5	6	7	8	9	10
テンポが適切である	1	2	3	4	5	6	7	8	9	10
経験豊かな	1	2	3	4	5	6	7	8	9	10
かん高い	1	2	3	4	5	6	7	8	9	10
流暢な	1	2	3	4	5	6	7	8	9	10
ポーズの長さが適切である	1	2	3	4	5	6	7	8	9	10
よく練習できている	1	2	3	4	5	6	7	8	9	10
速い	1	2	3	4	5	6	7	8	9	10
分かりやすい	1	2	3	4	5	6	7	8	9	10
音質が粗い	1	2	3	4	5	6	7	8	9	10
親しみやすい	1	2	3	4	5	6	7	8	9	10

问卷调查

姓名:　　性别:　　年龄:

大学:　　专业:　　年级:

何时开始学习英语？

分别学习了多少年？

是否有英语，测验成绩， 分别多少分？ TOEIC， IELTS， TOFEL，N2

如何学习的英语？

在大学时是否有英语课，如果有英语课是什么样的英语课？

在日常生活中是否有机会使用英语？如果有是什么样的场合？

对于自己现阶段的英语水平，有什么在意的点么？想要提高哪方面的能力？

是否受过英语公共演讲相关训练？如果有，是什么样的训练？

对于今天的实验，有何感想？

Appendix B. Sound pressure level (SPL) of stimuli in Experiment 1 (Fast peak)

Stimuli Number	1st (dB LAF)	2nd (dB LAF)	3rd (dB LAF)	Avg (dB LAF)
Pure Tone	62.0	62.0	62.0	62.0
0	67.1	67.1	67.1	67.1
1	67.1	67.1	67.1	67.1
2	67.1	67.1	67.1	67.1
3	67.1	67.1	67.1	67.1
4	67.1	67.1	67.1	67.1
5	67.1	67.1	67.1	67.1
6	67.1	67.1	67.1	67.1
7	66.5	66.5	66.5	66.5
8	66.5	66.5	66.5	66.5
9	66.5	66.5	66.5	66.5
10	66.5	66.5	66.5	66.5
11	66.5	66.5	66.5	66.5
12	66.5	66.5	66.5	66.5
13	66.5	66.5	66.5	66.5
14	68.7	68.7	68.7	68.7
15	68.7	68.7	68.7	68.7
16	68.7	68.7	68.7	68.7
17	68.7	68.7	68.7	68.7
18	68.7	68.7	68.7	68.7
19	68.7	68.7	68.7	68.7
20	68.7	68.7	68.7	68.7

Appendix C. Factor analysis of Experiment 1 data

English group

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.947
Bartlett's Test of Sphericity	Approx. Chi-Square	8159.853
	df	253
	Sig.	0.000

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared			Rotation Sums of Squared		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	11.259	48.953	48.953	11.259	48.953	48.953	9.925	43.153	43.153
2	2.569	11.171	60.124	2.569	11.171	60.124	3.829	16.646	59.799
3	1.665	7.239	67.363	1.665	7.239	67.363	1.656	7.200	67.000
4	1.025	4.457	71.820	1.025	4.457	71.820	1.109	4.821	71.820
5	0.969	4.212	76.032						
6	0.838	3.642	79.674						
7	0.574	2.494	82.169						
8	0.506	2.199	84.367						
9	0.475	2.063	86.431						
10	0.427	1.855	88.285						
11	0.390	1.698	89.983						
12	0.342	1.486	91.469						
13	0.306	1.331	92.800						
14	0.285	1.240	94.041						
15	0.215	0.935	94.975						
16	0.205	0.892	95.867						
17	0.189	0.820	96.687						
18	0.160	0.694	97.381						
19	0.156	0.679	98.060						
20	0.144	0.625	98.685						
21	0.119	0.518	99.203						
22	0.106	0.461	99.664						
23	0.077	0.336	100.000						

Extraction Method: Principal Component Analysis.

Rotated Component Matrix[a]

	Component			
	1	2	3	4
elegant	0.859	0.154	-0.022	-0.051
skillful	0.821	0.359	0.026	-0.070
smooth	0.818	0.280	-0.076	-0.006
with appropriate rhythm	0.813	0.266	0.104	0.023
natural	0.811	0.333	0.035	-0.037
experienced	0.787	0.402	-0.056	-0.090
well-practiced	0.781	0.415	0.018	-0.076
with appropriate pause duration	0.779	0.340	0.153	0.022
clear-cut	0.776	-0.072	-0.054	0.149
dynamic	0.771	0.079	0.065	0.122
at a suitable tempo	0.771	0.378	0.096	0.036
polite	0.736	-0.014	0.064	-0.233
friendly	0.728	0.086	0.124	-0.161
fluent	0.713	0.460	-0.083	-0.061
intelligible	0.699	0.054	-0.199	0.265
easy to understand	0.674	-0.096	-0.224	0.355
nervous	-0.487	0.165	0.409	-0.024
speedy	0.170	0.923	0.002	0.036
rushed	0.201	0.923	0.041	0.057
fast	0.222	0.908	0.021	0.039
high-pitched	0.025	0.069	0.841	0.028
shrill	0.115	-0.089	0.729	0.179
rough-timbred	-0.065	0.113	0.264	0.853

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 6 iterations.

Chinese group

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.944
Bartlett's Test of Sphericity	Approx. Chi-Square	8701.652
	df	253
	Sig.	0.000

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared			Rotation Sums of Squared		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	11.186	48.633	48.633	11.186	48.633	48.633	10.036	43.633	43.633
2	3.309	14.386	63.019	3.309	14.386	63.019	3.411	14.832	58.465
3	1.365	5.936	68.954	1.365	5.936	68.954	1.925	8.371	66.836
4	1.087	4.728	73.682	1.087	4.728	73.682	1.575	6.846	73.682
5	0.798	3.470	77.153						
6	0.730	3.174	80.327						
7	0.630	2.738	83.065						
8	0.499	2.170	85.235						
9	0.443	1.928	87.163						
10	0.369	1.605	88.768						
11	0.305	1.327	90.095						
12	0.293	1.276	91.371						
13	0.258	1.123	92.495						
14	0.253	1.098	93.593						
15	0.219	0.953	94.546						
16	0.213	0.927	95.473						
17	0.209	0.907	96.380						
18	0.170	0.741	97.121						
19	0.162	0.706	97.827						
20	0.140	0.610	98.437						
21	0.129	0.562	99.000						
22	0.122	0.530	99.530						
23	0.108	0.470	100.000						

Extraction Method: Principal Component Analysis.

Rotated Component Matrix[a]

	Component			
	1	2	3	4
自然的	0.902	0.123	0.075	-0.004
节奏适当的	0.880	0.130	0.039	-0.052
经验丰富的	0.871	0.188	0.134	0.057
节拍适当的	0.869	0.192	0.027	-0.023
停顿时长适当的	0.864	0.164	0.050	-0.042
擅长的	0.832	0.265	0.179	0.026
充分练习的	0.787	0.388	0.172	-0.026
礼貌的	0.776	-0.035	0.237	-0.058
友好的	0.766	-0.123	0.131	-0.096
优雅的	0.736	-0.045	0.400	0.079
明白易懂的	0.714	-0.249	0.352	0.047
容易理解的	0.702	-0.258	0.312	0.071
流畅的	0.691	0.501	0.260	0.048
流利的	0.685	0.526	0.190	0.028
动感的	0.651	0.453	0.067	0.134
紧张的	-0.504	0.305	-0.275	0.048
语速快的	0.071	0.883	-0.019	0.147
匆忙的	-0.143	0.841	-0.079	0.149
有速度感的	0.477	0.738	0.003	0.038
音质粗糙的	-0.101	-0.003	-0.830	0.229
口齿清脆的	0.470	0.037	0.703	0.127
声调高的	0.109	0.072	-0.039	0.871
尖锐的	-0.143	0.221	-0.074	0.803

Extraction Method: Principal Component

a. Rotation converged in 5 iterations.

Japanese group

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.934
Bartlett's Test of Sphericity	Approx. Chi-Square	7610.259
	df	253
	Sig.	0.000

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared			Rotation Sums of Squared		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	10.410	45.260	45.260	10.410	45.260	45.260	8.703	37.838	37.838
2	3.262	14.184	59.444	3.262	14.184	59.444	4.756	20.677	58.515
3	1.609	6.996	66.440	1.609	6.996	66.440	1.800	7.825	66.340
4	1.071	4.658	71.097	1.071	4.658	71.097	1.094	4.757	71.097
5	0.870	3.784	74.881						
6	0.749	3.255	78.136						
7	0.663	2.883	81.019						
8	0.522	2.272	83.291						
9	0.493	2.145	85.436						
10	0.466	2.028	87.463						
11	0.385	1.674	89.137						
12	0.364	1.585	90.722						
13	0.294	1.278	92.000						
14	0.285	1.238	93.238						
15	0.246	1.071	94.309						
16	0.239	1.037	95.346						
17	0.219	0.953	96.299						
18	0.185	0.803	97.102						
19	0.162	0.706	97.808						
20	0.158	0.685	98.494						
21	0.134	0.582	99.076						
22	0.114	0.497	99.573						
23	0.098	0.427	100.000						

Extraction Method: Principal Component Analysis.

Rotated Component Matrix[a]

	Component			
	1	2	3	4
テンポが適切である	0.886	0.229	-0.004	0.038
ポーズの長さが適切である	0.875	0.253	-0.027	0.034
リズムが適切である	0.872	0.211	0.041	0.024
分かりやすい	0.790	-0.021	0.147	-0.045
自然な	0.780	0.380	0.004	0.004
上手な	0.776	0.286	0.101	-0.095
よく練習できている	0.756	0.369	-0.083	-0.115
親しみやすい	0.742	-0.059	0.055	0.149
聴き取りやすい	0.725	-0.115	0.236	0.003
丁寧な	0.718	-0.404	0.187	0.004
経験豊か	0.689	0.317	-0.083	-0.192
滑らかな	0.643	0.551	0.092	-0.156
優雅な	0.627	0.082	0.330	-0.122
流暢な	0.618	0.564	0.039	-0.149
歯切れの良い	0.614	0.508	0.153	0.016
速い	0.113	0.898	0.031	-0.007
スピード感がある	0.276	0.877	0.007	-0.063
せきたてるような	-0.061	0.851	0.133	0.016
緊張感がある	0.081	0.675	0.187	0.215
活力のある	0.481	0.497	0.188	0.128
高音である	0.127	0.086	0.856	0.004
かん高い	0.075	0.216	0.833	-0.019
音質が粗い	-0.046	0.094	-0.029	0.937

Extraction Method: Principal Component

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 5 iterations.

Appendix D. Statistical analysis of Experiment 1 data

・Speech rate factor

English group

Normality test of Speech Rate factor

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
0.075	0.185	19	0.088	0.915	19	0.090
0.15	0.125	19	.200*	0.975	19	0.862
0.3	0.087	19	.200*	0.981	19	0.949
0.6	0.176	19	0.125	0.884	19	0.025
1.2	0.118	19	.200*	0.969	19	0.760
2.4	0.085	19	.200*	0.978	19	0.920
4.8	0.155	19	.200*	0.915	19	0.090

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Friedman test between manipulated pause duration stimulus types

Test Statistics[a]

N	19
Chi-Square	104.414
df	6
Asymp. Sig.	0.000

a. Friedman Test

Holm-Bonferroni test of speech rate factor

Pair		P value	α Value	Result
0.30	0.075	0.000	0.002381	significant
0.60	0.075	0.000	0.0025	significant
1.20	0.075	0.000	0.002632	significant
2.40	0.075	0.000	0.002778	significant
4.80	0.075	0.000	0.002941	significant
0.60	0.15	0.000	0.003125	significant
1.20	0.15	0.000	0.003333	significant
2.40	0.15	0.000	0.003571	significant
4.80	0.15	0.000	0.003846	significant
0.60	0.30	0.000	0.004167	significant
1.20	0.30	0.000	0.004545	significant
2.40	0.30	0.000	0.005	significant
4.80	0.30	0.000	0.005556	significant
2.40	0.60	0.000	0.00625	significant
0.30	0.15	0.000	0.007143	significant
4.80	0.60	0.000	0.008333	significant
1.20	0.60	0.000	0.01	significant
2.40	1.20	0.001	0.0125	significant
0.15	0.075	0.002	0.016667	significant
4.80	1.20	0.004	0.025	significant
4.80	2.40	0.421	0.05	not significant

Chinese group
Normality test of Speech Rate factor

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
0.075	0.139	20	.200*	0.955	20	0.443
0.15	0.122	20	.200*	0.959	20	0.534
0.3	0.109	20	.200*	0.970	20	0.762
0.6	0.147	20	.200*	0.926	20	0.129
1.2	0.144	20	.200*	0.955	20	0.456
2.4	0.170	20	0.132	0.898	20	0.037
4.8	0.125	20	.200*	0.978	20	0.902

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Friedman test between manipulated pause duration stimulus types

Test Statistics[a]

N	20
Chi-Square	111.343
df	6
Asymp. Sig.	0.000

a. Friedman Test

Holm-Bonferroni test of speech rate factor

Pairs		P value	α Value	Result
0.60	0.075	0.000	0.002380952	significant
1.20	0.075	0.000	0.0025	significant
2.40	0.075	0.000	0.002631579	significant
4.80	0.075	0.000	0.002777778	significant
1.20	0.15	0.000	0.002941176	significant
2.40	0.15	0.000	0.003125	significant
4.80	0.15	0.000	0.003333333	significant
2.40	0.30	0.000	0.003571429	significant
4.80	0.30	0.000	0.003846154	significant
2.40	0.60	0.000	0.004166667	significant
4.80	0.60	0.000	0.004545455	significant
4.80	1.20	0.000	0.005	significant
0.30	0.15	0.000	0.005555556	significant
0.60	0.15	0.000	0.00625	significant
1.20	0.30	0.000	0.007142857	significant
0.30	0.075	0.000	0.008333333	significant
2.40	1.20	0.000	0.01	significant
0.60	0.30	0.000	0.0125	significant
1.20	0.60	0.001	0.016666667	significant
4.80	2.40	0.008	0.025	significant
0.15	0.075	0.048	0.05	significant

Japanese group
Normality test of Speech Rate factor

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
0.075	0.145	19	.200*	0.930	19	0.171
0.15	0.107	19	.200*	0.964	19	0.646
0.3	0.124	19	.200*	0.939	19	0.249
0.6	0.163	19	.200*	0.944	19	0.305
1.2	0.153	19	.200*	0.929	19	0.164
2.4	0.098	19	.200*	0.986	19	0.991
4.8	0.138	19	.200*	0.937	19	0.229

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Friedman test between manipulated pause duration stimulus types

Test Statistics[a]

N	19
Chi-Square	97.263
df	6
Asymp. Sig.	0.000

a. Friedman Test

Holm-Bonferroni test of speech rate factor

Pairs		P value	α Value	Result
2.4	0.075	0.000	0.002381	significant
4.8	0.075	0.000	0.0025	significant
1.2	0.15	0.000	0.002632	significant
2.4	0.15	0.000	0.002778	significant
4.8	0.15	0.000	0.002941	significant
2.4	0.3	0.000	0.003125	significant
4.8	0.3	0.000	0.003333	significant
2.4	0.6	0.000	0.003571	significant
4.8	0.6	0.000	0.003846	significant
1.2	0.075	0.000	0.004167	significant
1.2	0.3	0.000	0.004545	significant
4.8	1.2	0.000	0.005	significant
0.6	0.15	0.000	0.005556	significant
0.3	0.075	0.000	0.00625	significant
0.6	0.075	0.000	0.007143	significant
0.3	0.15	0.001	0.008333	significant
1.2	0.6	0.001	0.01	significant
2.4	1.2	0.002	0.0125	significant
0.6	0.3	0.007	0.016667	significant
4.8	2.4	0.022	0.025	significant
0.15	0.075	0.243	0.05	not significant

・Naturalness factor

English group

Normality test of Naturalness factor

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
0.075	0.160	19	.200*	0.887	19	0.029
0.15	0.155	19	.200*	0.895	19	0.039
0.3	0.170	19	0.152	0.933	19	0.200
0.6	0.127	19	.200*	0.947	19	0.348
1.2	0.100	19	.200*	0.963	19	0.632
2.4	0.087	19	.200*	0.971	19	0.795
4.8	0.130	19	.200*	0.978	19	0.912

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Friedman test between manipulated pause duration stimulus types

Test Statistics[a]

N	19
Chi-Square	93.293
df	6
Asymp. Sig.	0.000

a. Friedman Test

Holm-Bonferroni test of Naturalness factor

Pair		P value	α Value	Result
0.15	0.075	0.000	0.002381	significant
0.30	0.075	0.000	0.0025	significant
0.60	0.075	0.000	0.002632	significant
0.60	0.15	0.000	0.002778	significant
4.80	0.15	0.000	0.002941	significant
4.80	0.30	0.000	0.003125	significant
2.40	0.60	0.000	0.003333	significant
4.80	0.60	0.000	0.003571	significant
2.40	0.30	0.000	0.003846	significant
4.80	1.20	0.000	0.004167	significant
0.60	0.30	0.000	0.004545	significant
0.30	0.15	0.000	0.005	significant
1.20	0.60	0.000	0.005556	significant
2.40	1.20	0.000	0.00625	significant
4.80	2.40	0.000	0.007143	significant
2.40	0.15	0.001	0.008333	significant
1.20	0.30	0.001	0.01	significant
4.80	0.075	0.003	0.0125	significant
1.20	0.075	0.011	0.016667	significant
1.20	0.15	0.629	0.025	not significant
2.40	0.075	0.904	0.05	not significant

Chinese group
Normality test of Naturalness factor

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
0.075	0.128	20	.200*	0.965	20	0.652
0.15	0.106	20	.200*	0.976	20	0.878
0.3	0.158	20	.200*	0.943	20	0.268
0.6	0.106	20	.200*	0.969	20	0.731
1.2	0.107	20	.200*	0.988	20	0.993
2.4	0.085	20	.200*	0.991	20	0.999
4.8	0.149	20	.200*	0.929	20	0.149

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Friedman test between manipulated pause duration stimulus types

Test Statistics[a]

N	20
Chi-Square	92.529
df	6
Asymp. Sig.	0.000

a. Friedman Test

Holm-Bonferroni test of Naturalness factor

Pairs		P value	α Value	Result
0.30	0.075	0.000	0.002381	significant
0.60	0.075	0.000	0.0025	significant
0.30	0.15	0.000	0.002632	significant
2.40	0.30	0.000	0.002778	significant
4.80	0.30	0.000	0.002941	significant
2.40	0.60	0.000	0.003125	significant
4.80	0.60	0.000	0.003333	significant
2.40	1.20	0.000	0.003571	significant
4.80	1.20	0.000	0.003846	significant
0.60	0.15	0.000	0.004167	significant
1.20	0.075	0.000	0.004545	significant
4.80	0.15	0.001	0.005	significant
1.20	0.60	0.001	0.005556	significant
4.80	2.40	0.002	0.00625	significant
4.80	0.075	0.002	0.007143	significant
0.10	0.075	0.012	0.008333	not significant
2.40	0.15	0.023	0.01	not significant
1.20	0.15	0.048	0.0125	not significant
1.20	0.30	0.057	0.016667	not significant
0.60	0.30	0.179	0.025	not significant
2.40	0.075	0.247	0.05	not significant

Japanese group
Normality test of Naturalness factor

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
0.075	0.177	19	0.118	0.931	19	0.177
0.15	0.113	19	.200*	0.966	19	0.700
0.3	0.162	19	.200*	0.958	19	0.530
0.6	0.146	19	.200*	0.945	19	0.317
1.2	0.132	19	.200*	0.961	19	0.600
2.4	0.128	19	.200*	0.970	19	0.778
4.8	0.091	19	.200*	0.983	19	0.972

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Friedman test between manipulated pause duration stimulus types

Test Statistics[a]

N	19
Chi-Square	73.850
df	6
Asymp. Sig.	0.000

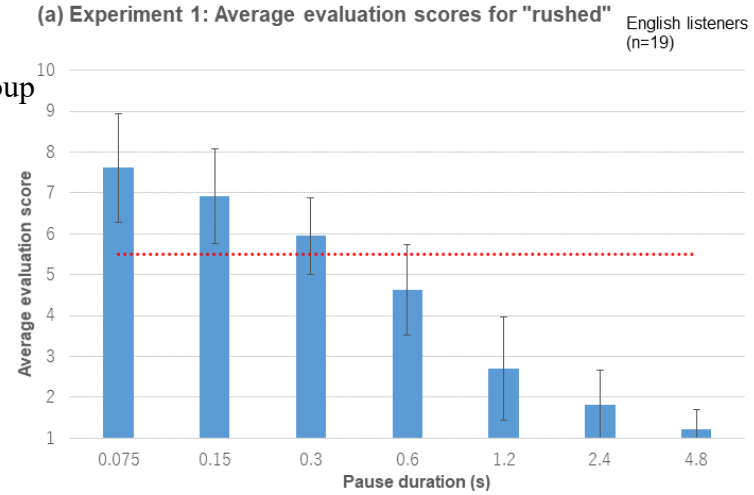
a. Friedman Test

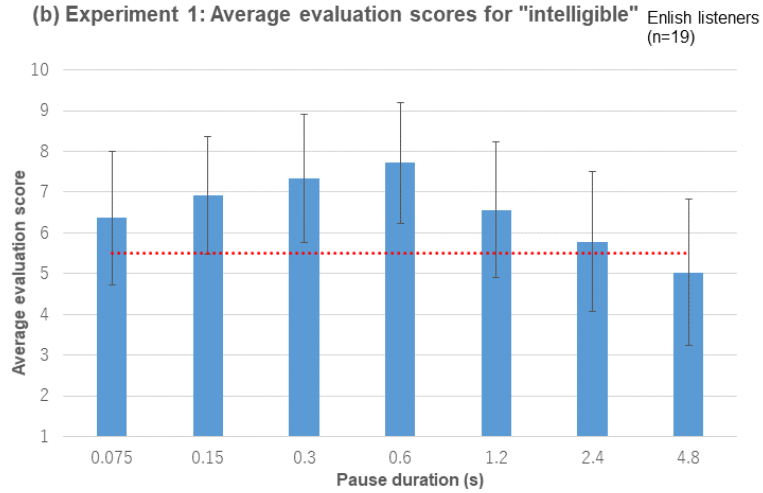
Holm-Bonferroni test of Naturalness factor

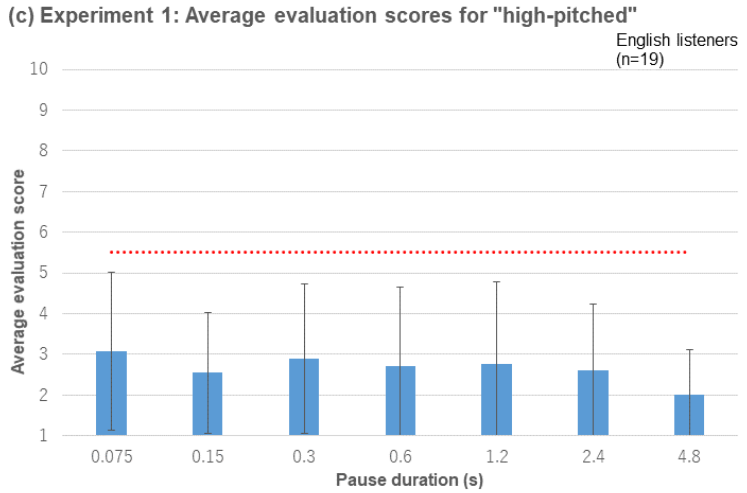
Pair		P value	α Value	Result
4.80	0.30	0.000	0.002381	significant
2.40	0.60	0.000	0.0025	significant
4.80	0.60	0.000	0.002632	significant
4.80	1.20	0.000	0.002778	significant
0.60	0.15	0.000	0.002941	significant
2.40	1.20	0.000	0.003125	significant
0.60	0.075	0.000	0.003333	significant
4.80	0.15	0.000	0.003571	significant
0.30	0.075	0.001	0.003846	significant
4.80	2.40	0.001	0.004167	significant
1.20	0.075	0.001	0.004545	significant
0.60	0.30	0.001	0.005	significant
2.40	0.30	0.001	0.005556	significant
4.80	0.075	0.002	0.00625	significant
0.30	0.15	0.004	0.007143	significant
1.20	0.15	0.005	0.008333	significant
0.15	0.075	0.107	0.01	not significant
1.20	0.60	0.107	0.0125	not significant
2.40	0.15	0.295	0.016667	not significant
1.20	0.30	0.295	0.025	not significant
2.40	0.075	0.573	0.05	not significant

Appendix E. Average evaluation scores for each evaluation item of Experiment 1

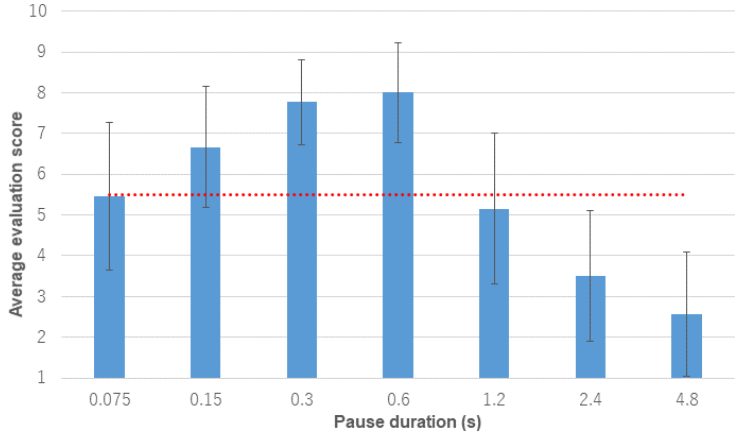
English group



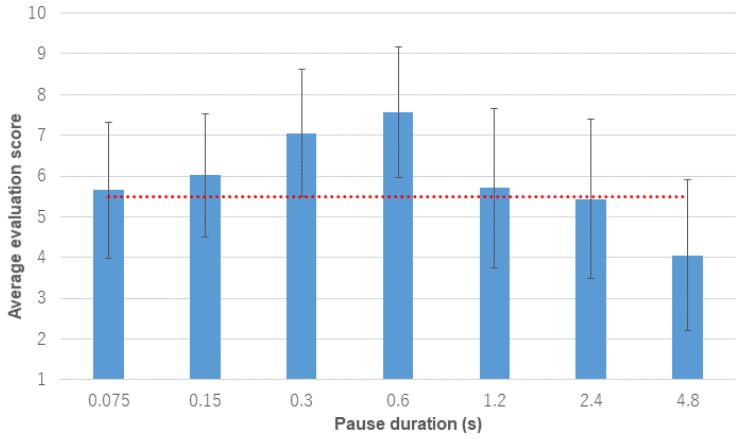




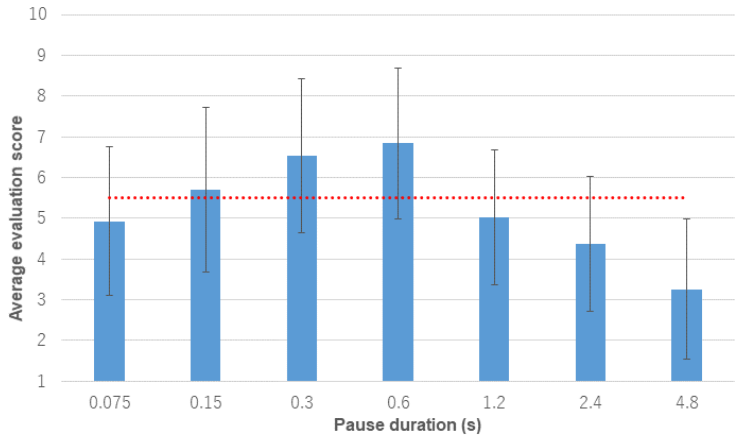
(d) Experiment 1: Average evaluation scores for "skillful" English listeners (n=19)



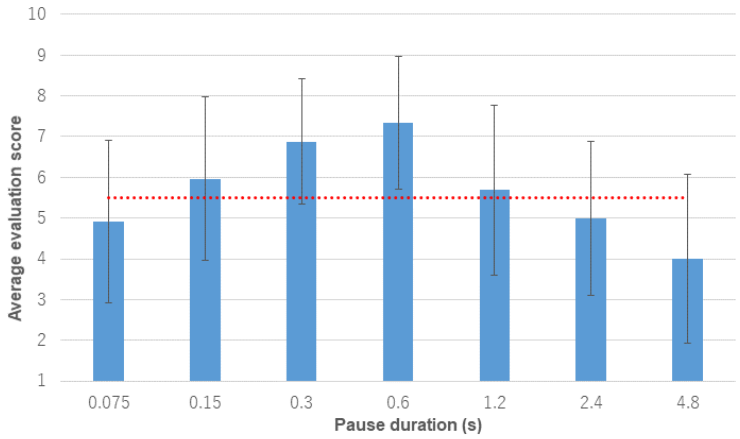
(e) Experiment 1: Average evaluation scores for "polite" English listeners (n=19)



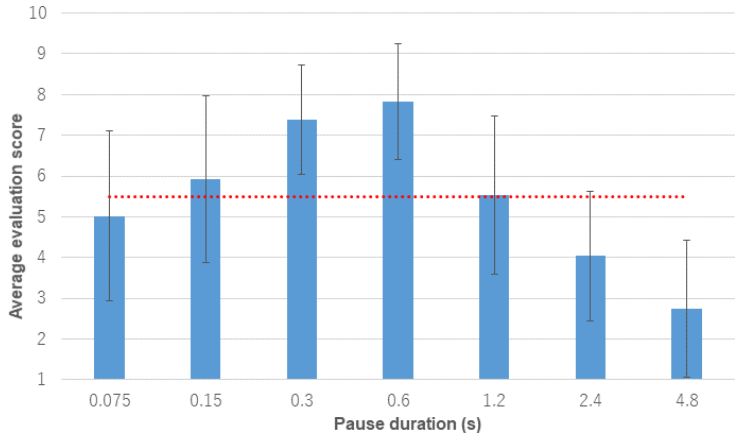
(f) Experiment 1: Average evaluation scores for "dynamic" English listeners (n=19)



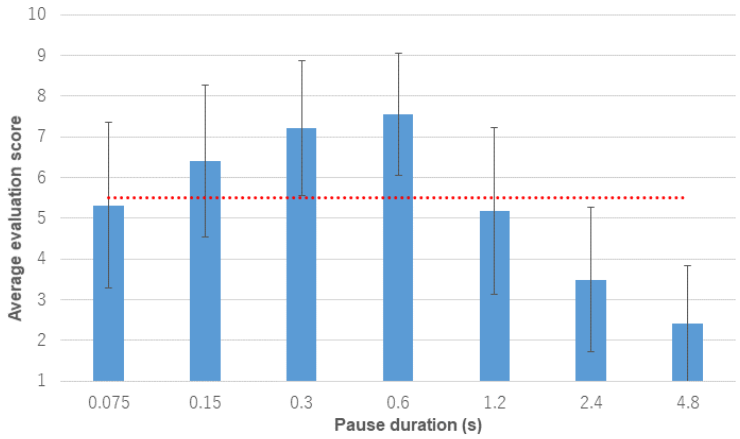
(g) Experiment 1: Average evaluation scores for "clear-cut" English listeners (n=19)



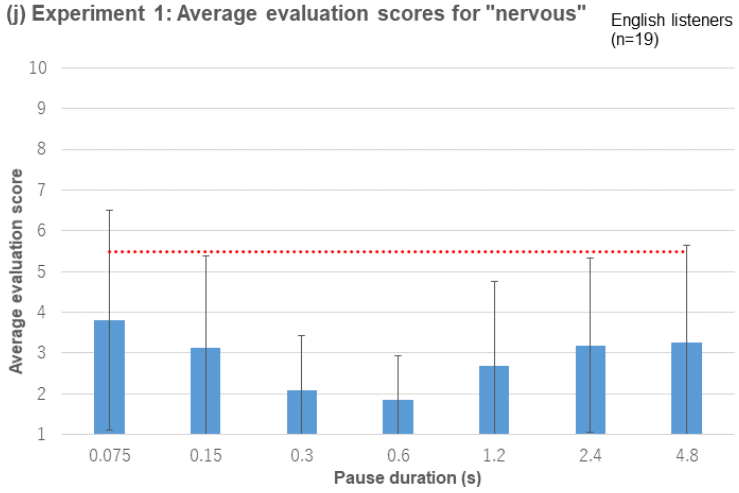
(h) Experiment 1: Average evaluation scores for "elegant" English listeners (n=19)



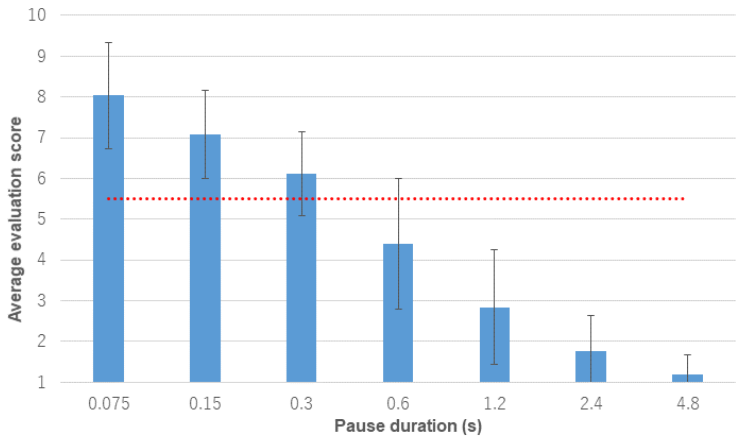
(i) Experiment 1: Average evaluation scores for "smooth" English listeners (n=19)



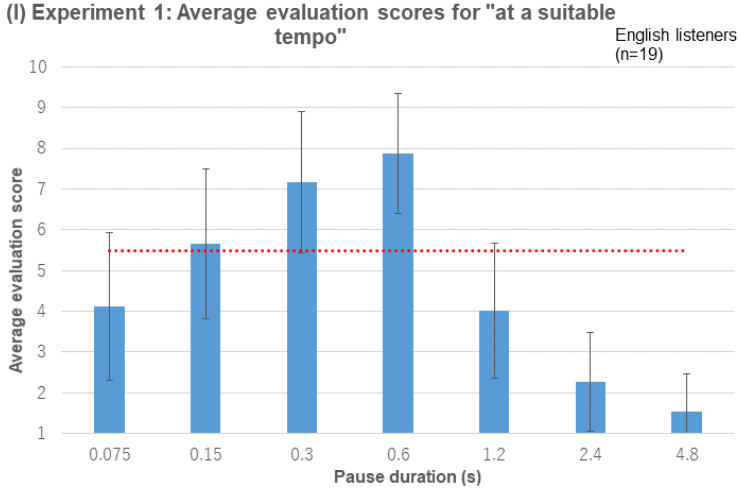
(j) Experiment 1: Average evaluation scores for "nervous"



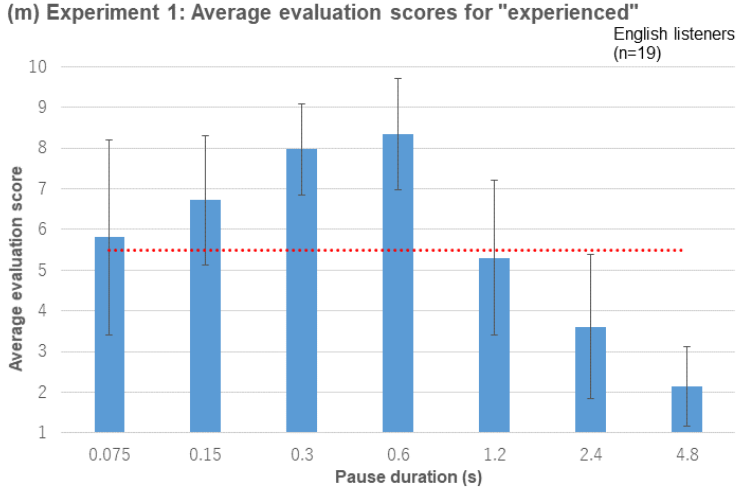
(k) Experiment 1: Average evaluation scores for "speedy"



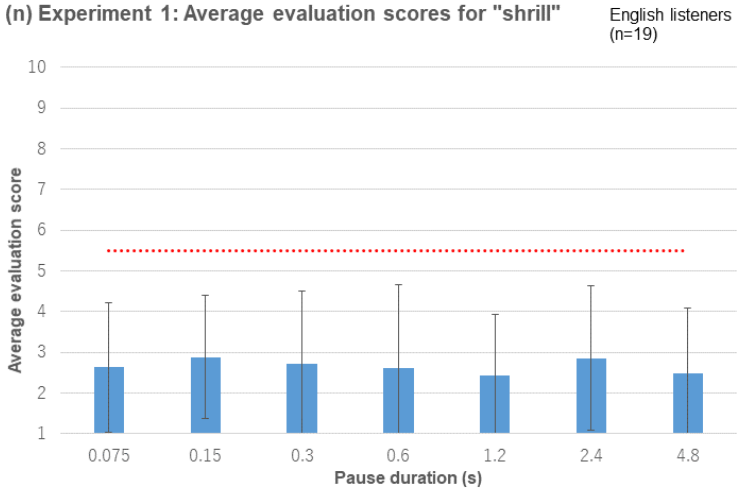
(l) Experiment 1: Average evaluation scores for "at a suitable tempo"



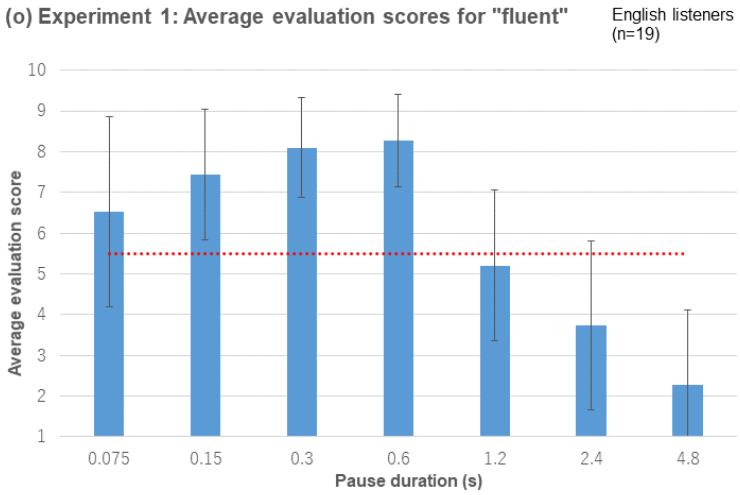
(m) Experiment 1: Average evaluation scores for "experienced"



(n) Experiment 1: Average evaluation scores for "shrill"

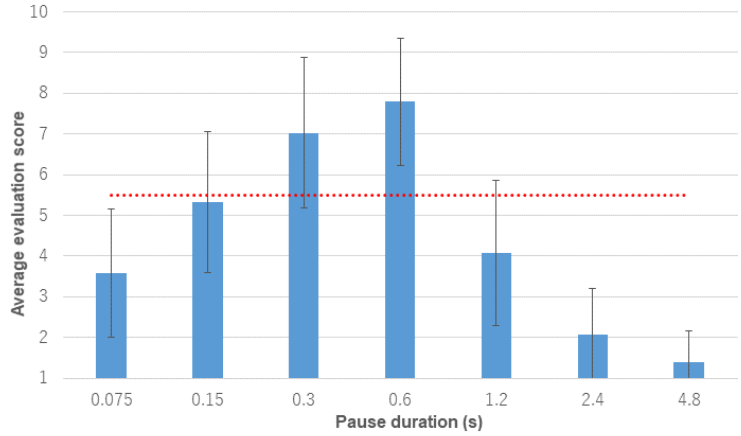


(o) Experiment 1: Average evaluation scores for "fluent"



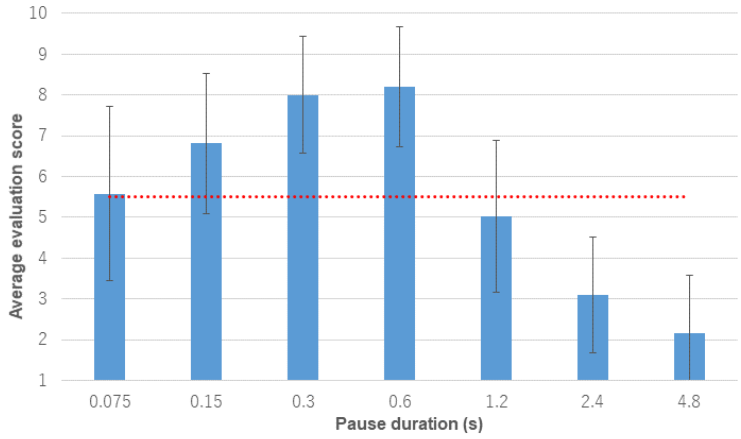
(p) Experiment 1: Average evaluation scores for "with appropriate pause duration"

English listeners (n=19)



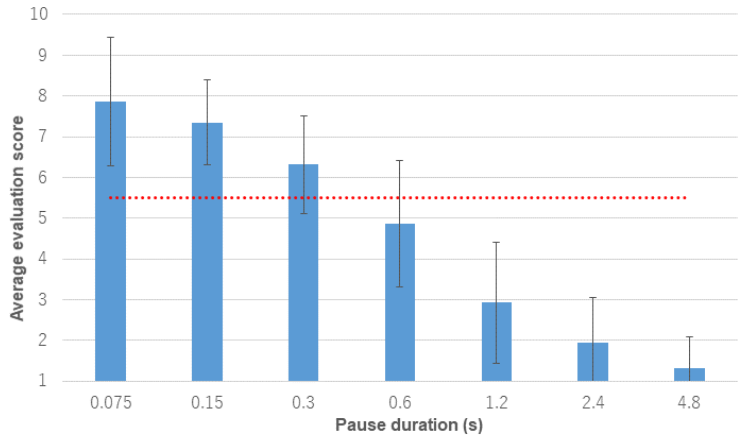
(q) Experiment 1: Average evaluation scores for "well-practiced"

English listeners (n=19)



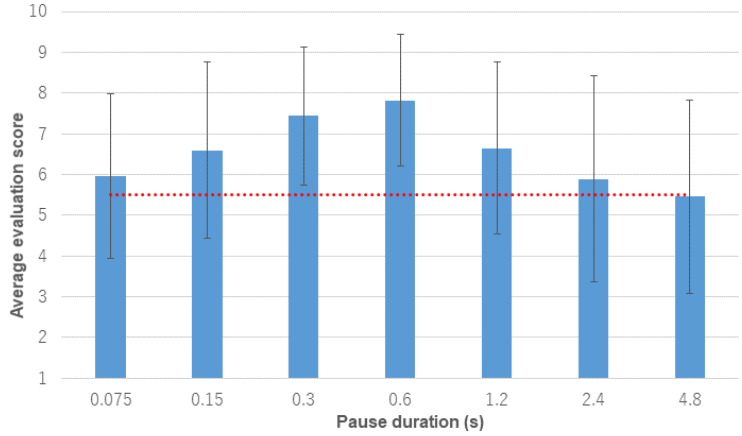
(r) Experiment 1: Average evaluation scores for "fast"

English listeners (n=19)



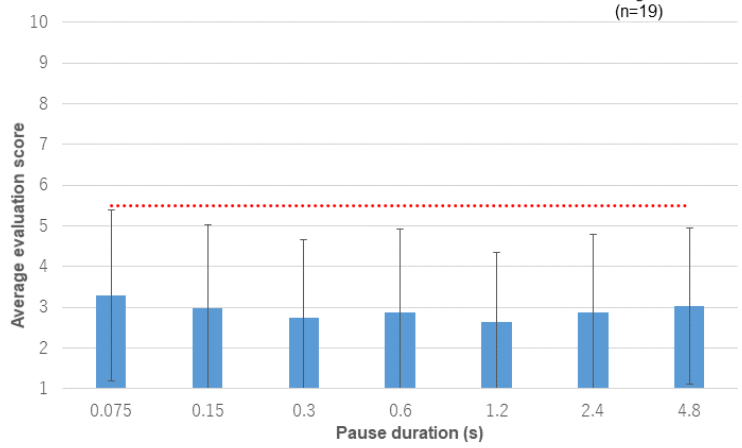
(s) Experiment 1: Average evaluation scores for "easy to understand"

English listeners (n=19)



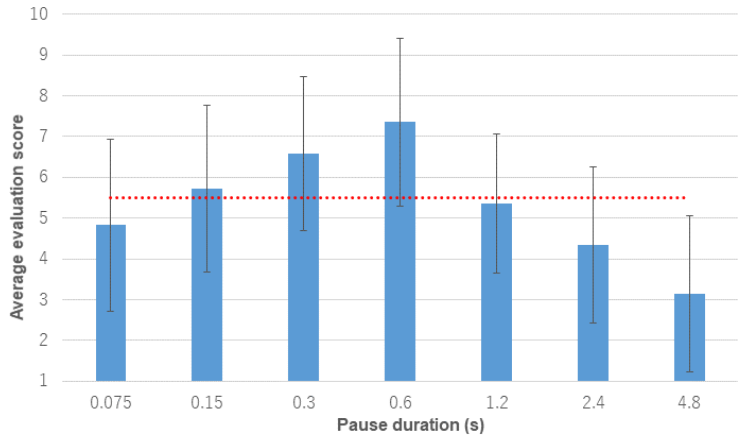
(t) Experiment 1: Average evaluation scores for "rough timbred"

English listeners (n=19)



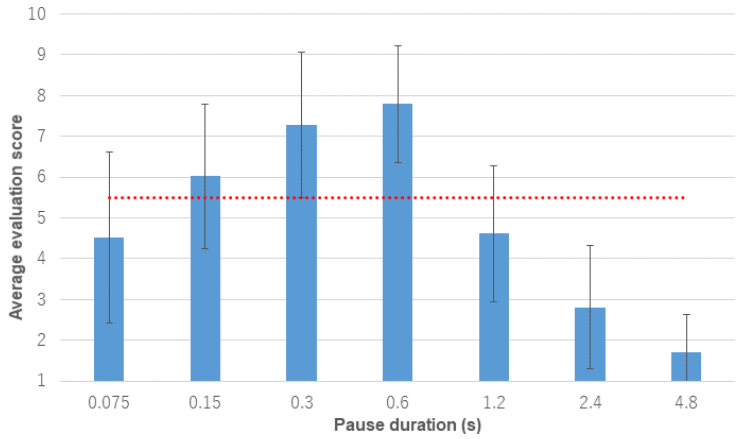
(u) Experiment 1: Average evaluation scores for "friendly"

English listeners (n=19)



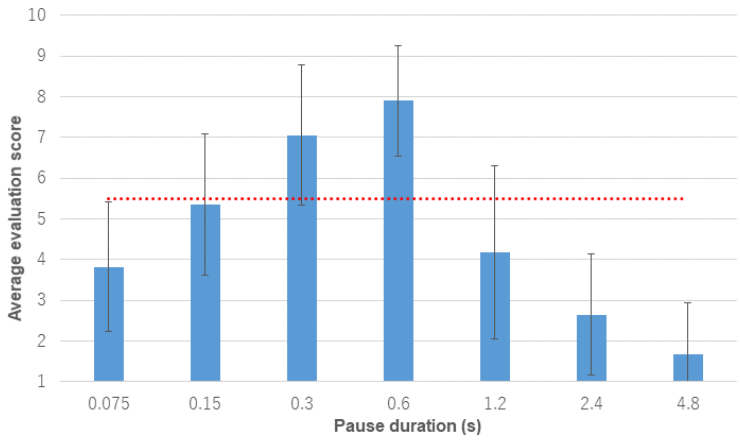
(v) Experiment 1: Average evaluation scores for "natural"

English listeners (n=19)

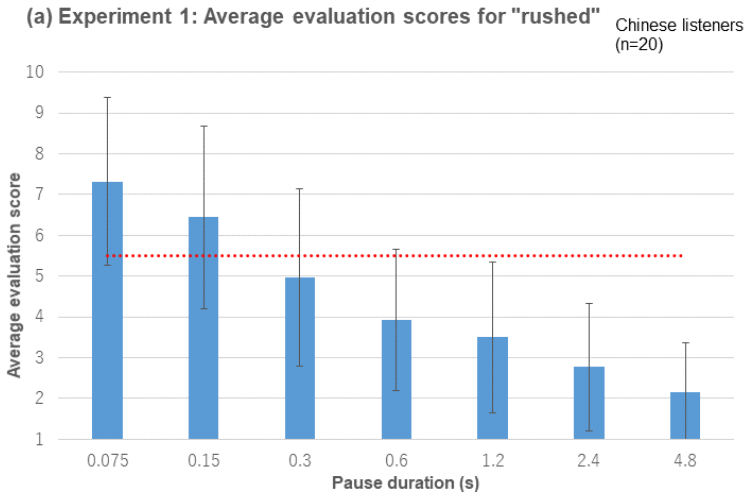


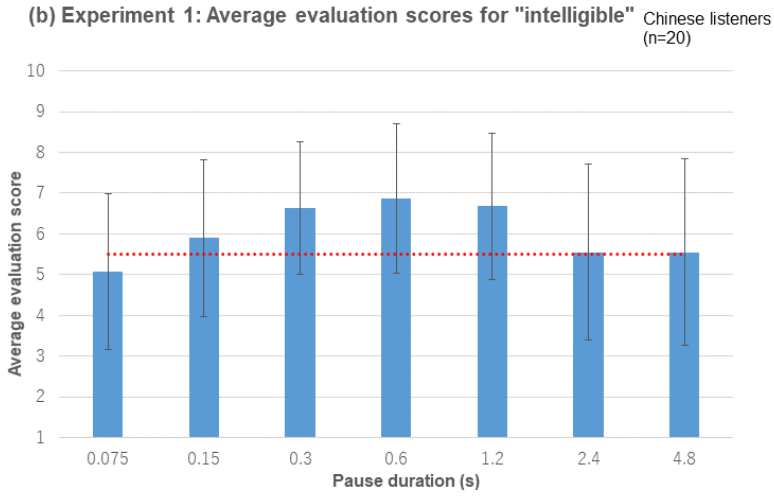
(w) Experiment 1: Average evaluation scores for "with appropriate ryhtem"

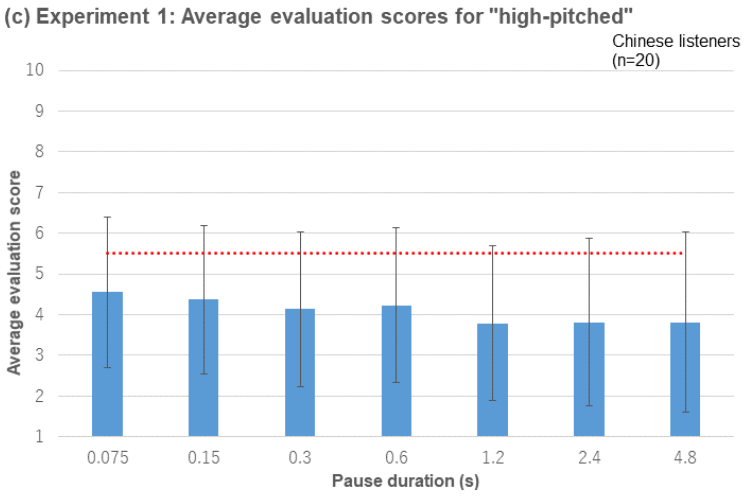
English listeners (n=19)



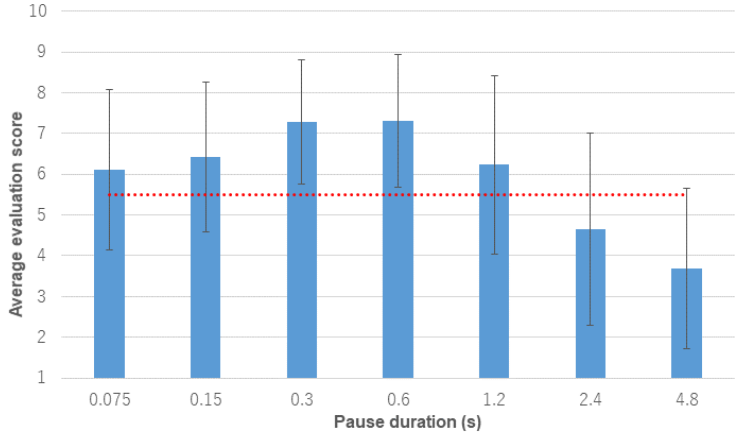
Chinese group



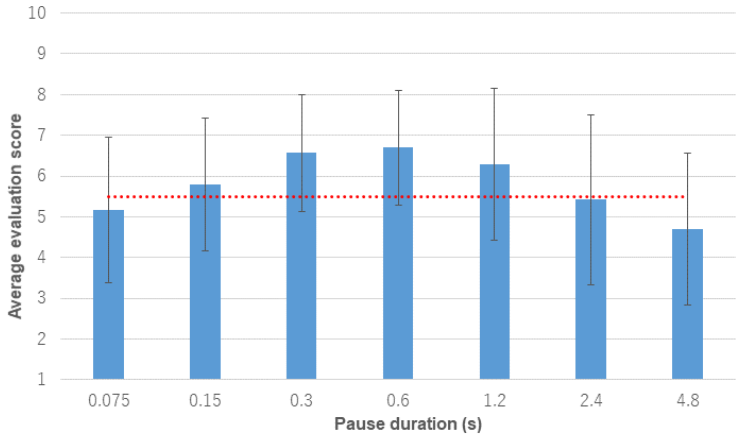




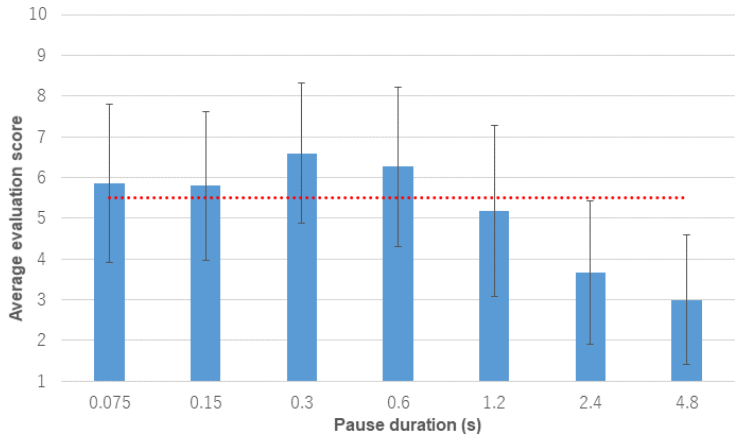
(d) Experiment 1: Average evaluation scores for "skillful" Chinese listeners (n=20)



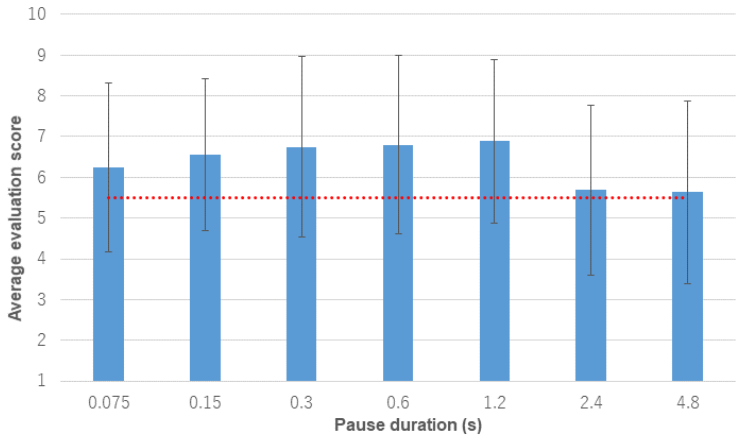
(e) Experiment 1: Average evaluation scores for "polite" Chinese listeners (n=20)



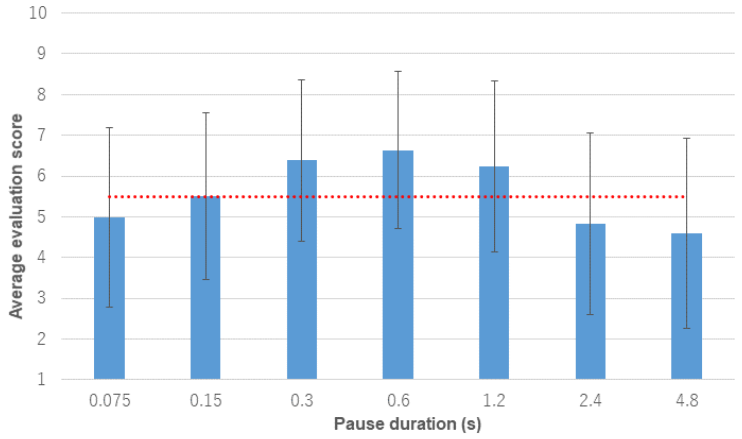
(f) Experiment 1: Average evaluation scores for "dynamic" Chinese listeners (n=20)



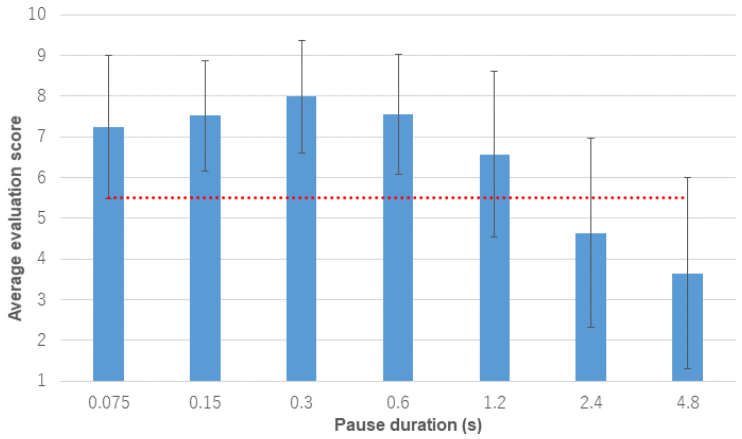
(g) Experiment 1: Average evaluation scores for "clear-cut" Chinese listeners (n=20)



(h) Experiment 1: Average evaluation scores for "elegant" Chinese listeners (n=20)

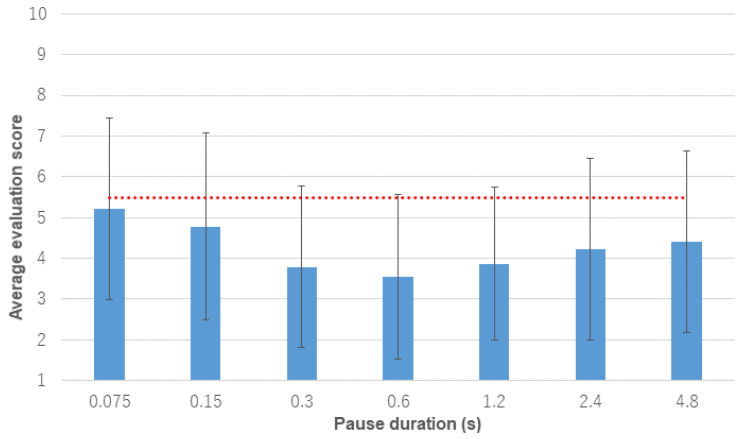


(i) Experiment 1: Average evaluation scores for "smooth" Chinese listeners (n=20)



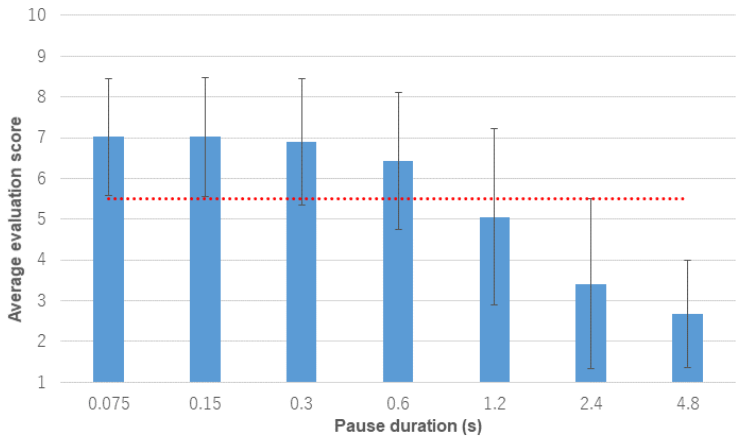
(j) Experiment 1: Average evaluation scores for "nervous"

Chinese listeners (n=20)



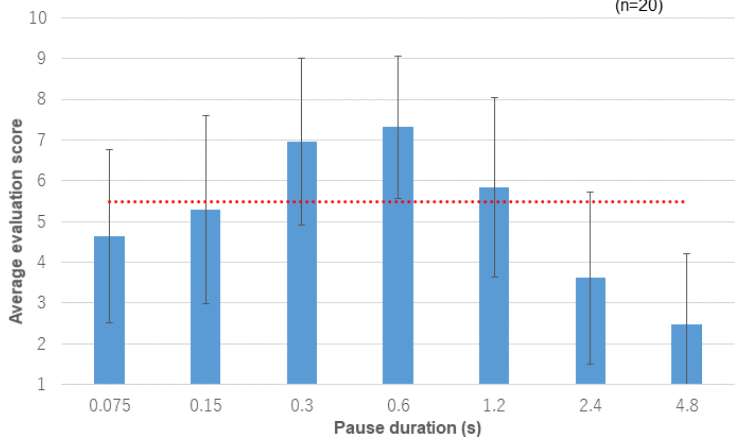
(k) Experiment 1: Average evaluation scores for "speedy"

Chinese listeners (n=20)

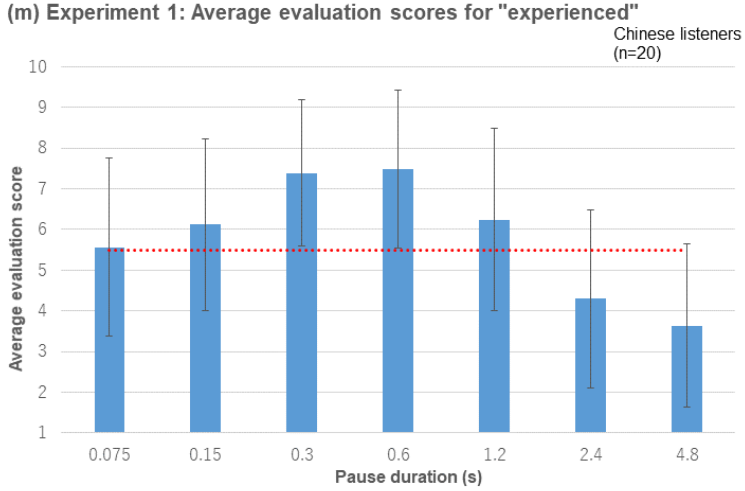


(l) Experiment 1: Average evaluation scores for "at a suitable tempo"

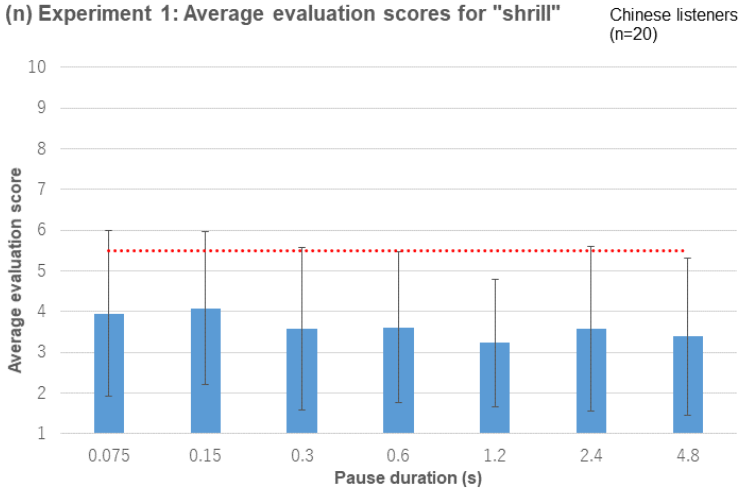
Chinese listeners (n=20)



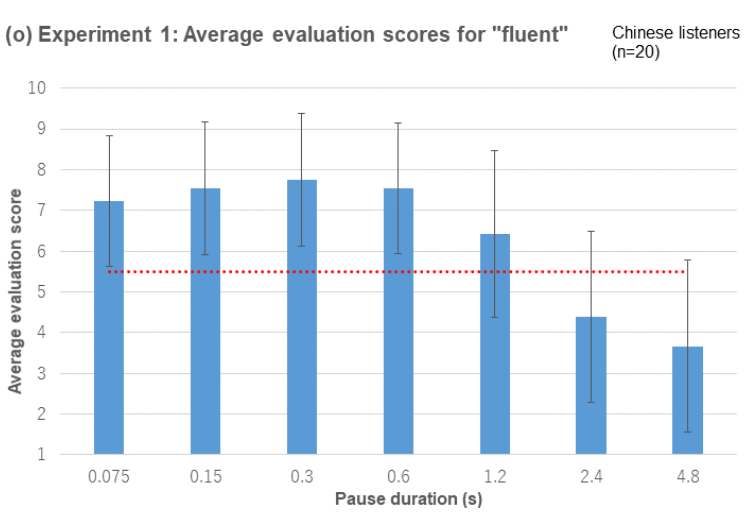
(m) Experiment 1: Average evaluation scores for "experienced"



(n) Experiment 1: Average evaluation scores for "shrill"

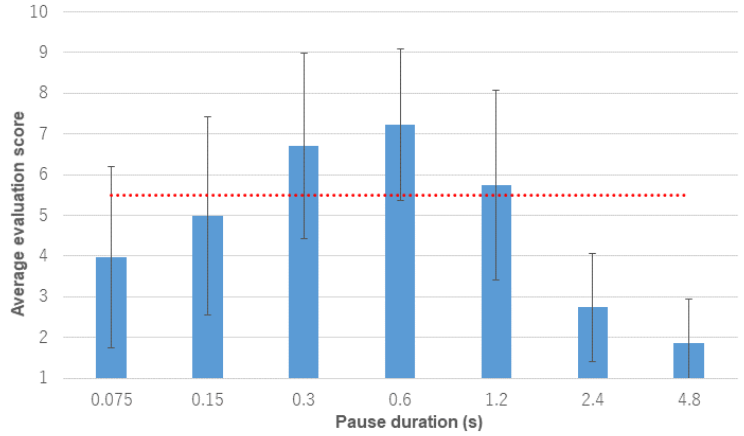


(o) Experiment 1: Average evaluation scores for "fluent"



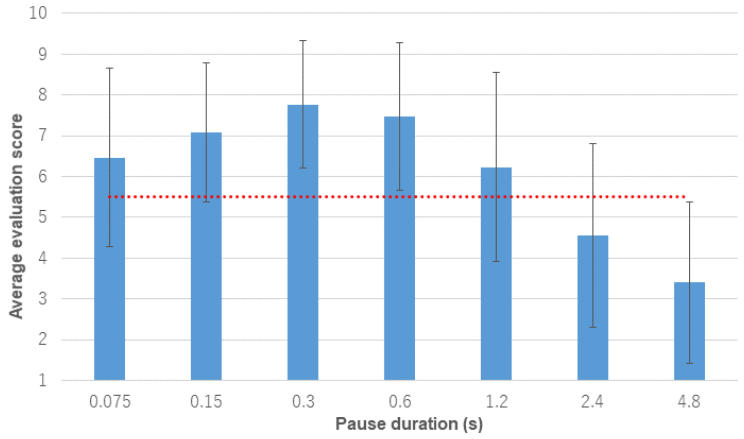
(p) Experiment 1: Average evaluation scores for "with appropriate pause duration"

Chinese listeners (n=20)



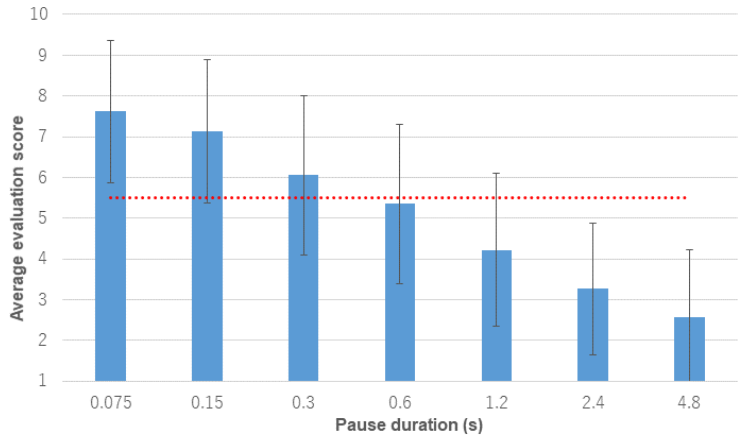
(q) Experiment 1: Average evaluation scores for "well-practiced"

CHinese listeners (n=20)

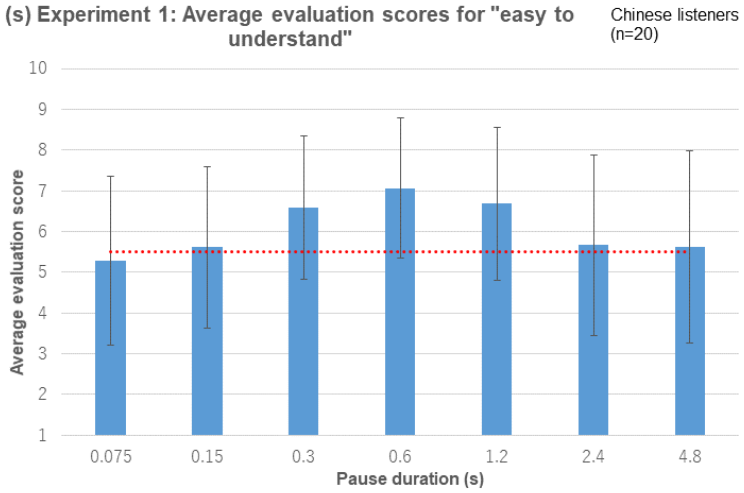


(r) Experiment 1: Average evaluation scores for "fast"

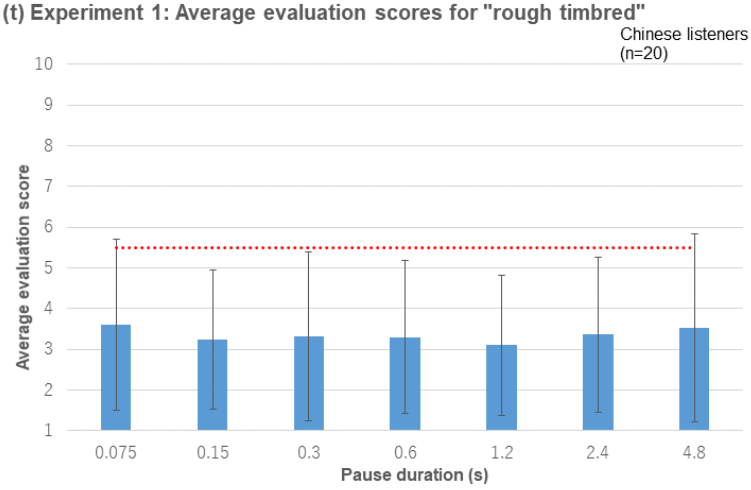
Chinese listeners (n=20)



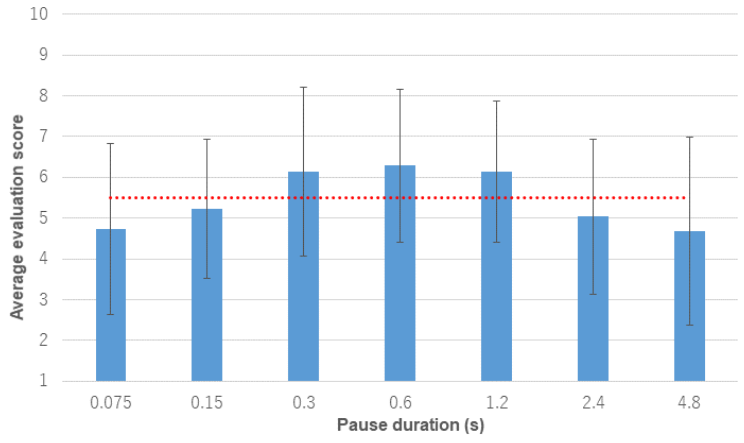
(s) Experiment 1: Average evaluation scores for "easy to understand"



(t) Experiment 1: Average evaluation scores for "rough timbred"

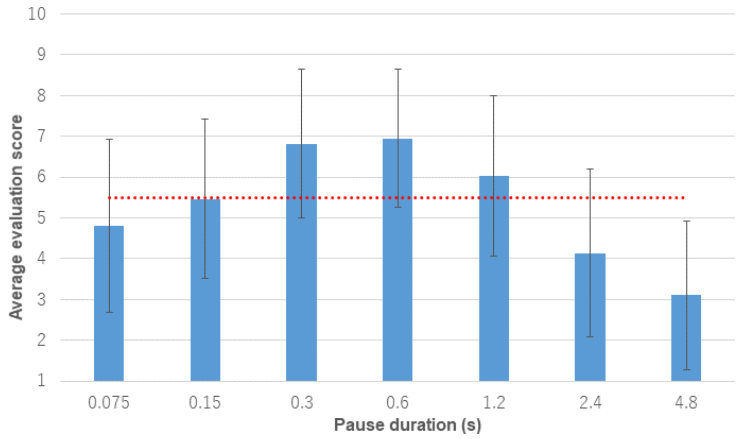


(u) Experiment 1: Average evaluation scores for "friendly"



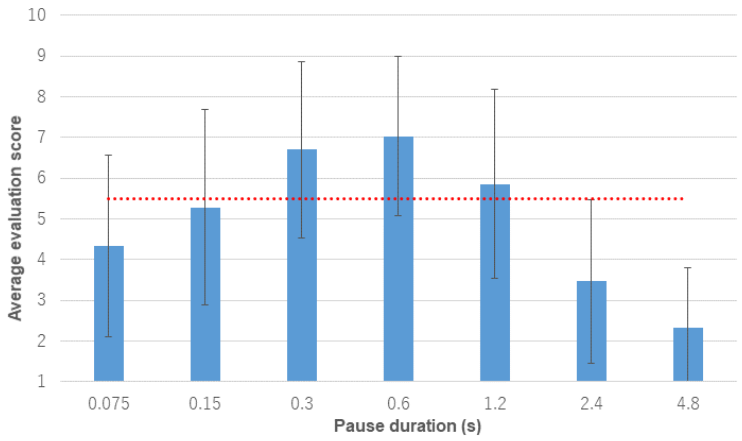
(v) Experiment 1: Average evaluation scores for "natural"

Chinese listeners (n=20)

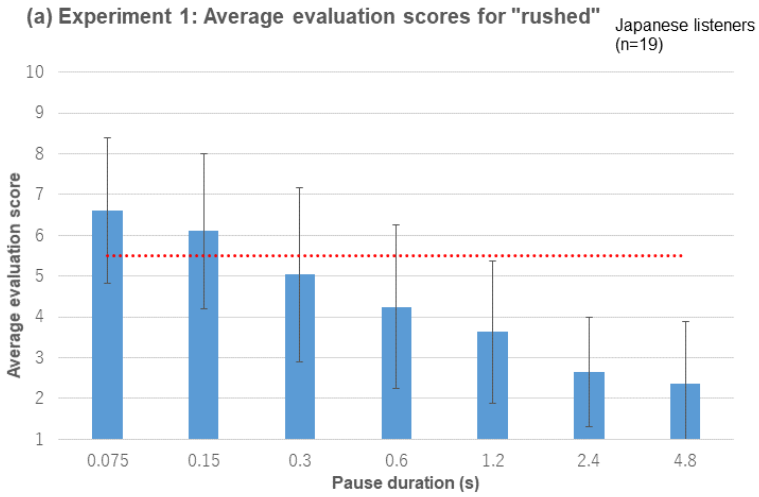


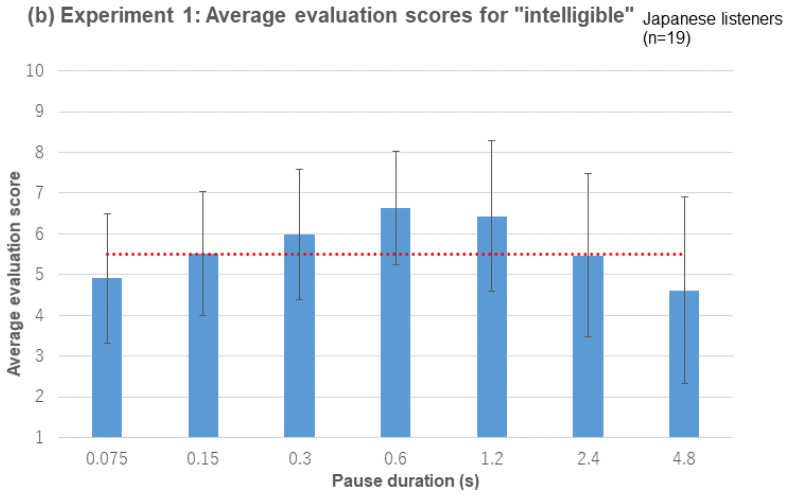
(w) Experiment 1: Average evaluation scores for "with appropriate ryhtem"

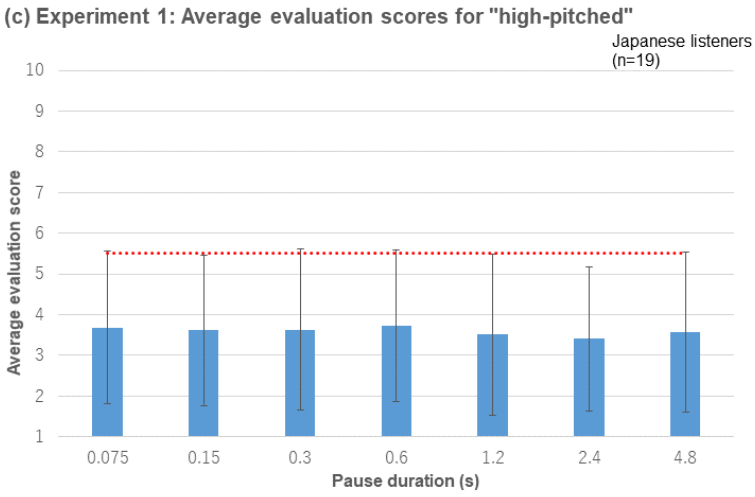
Chinese listeners (n=20)



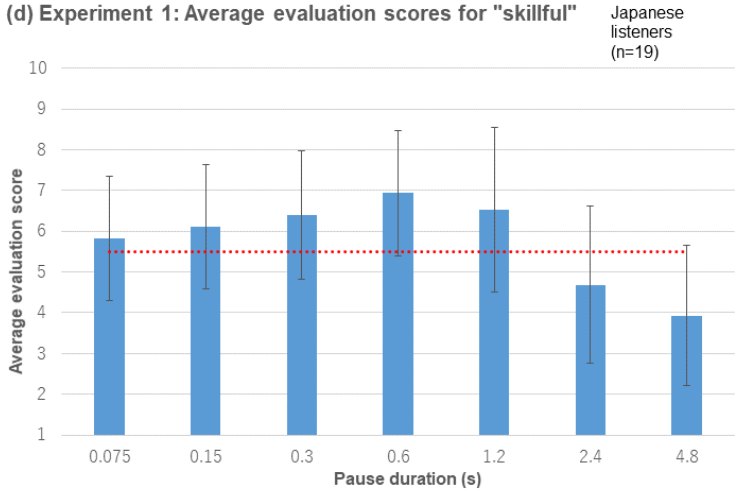
Japanese group



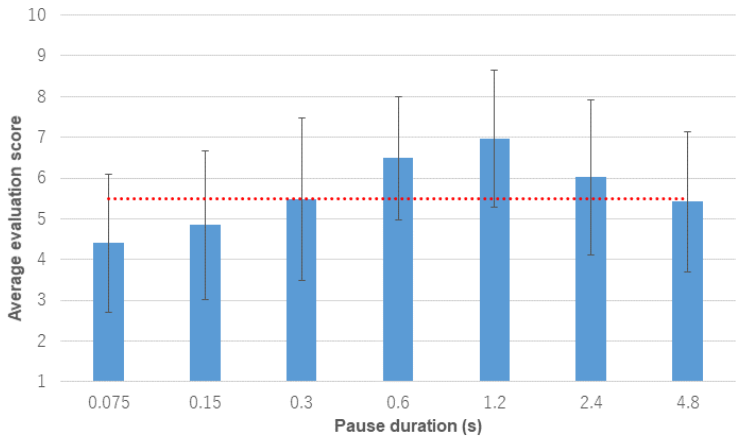




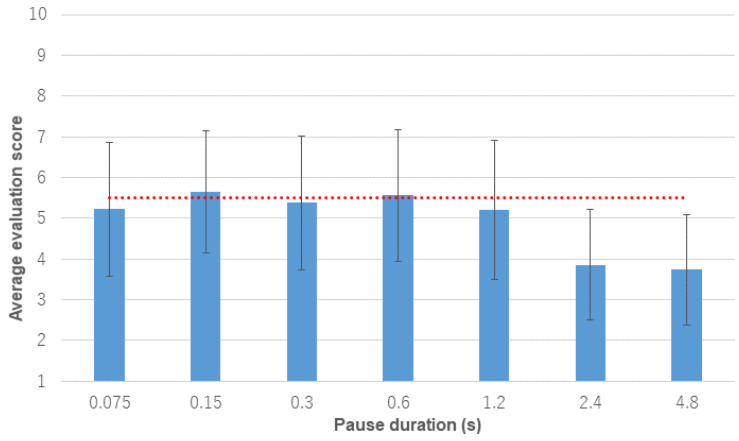
(d) Experiment 1: Average evaluation scores for "skillful"



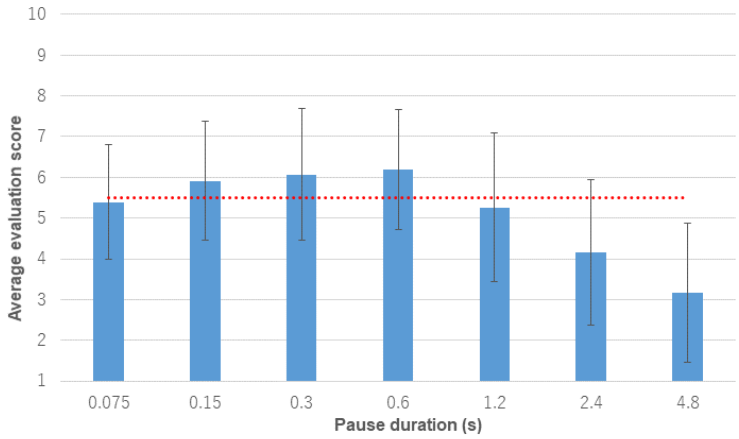
(e) Experiment 1: Average evaluation scores for "polite"



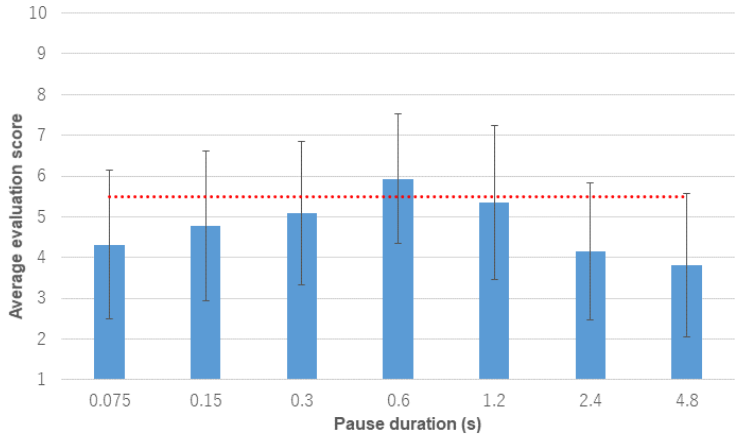
(f) Experiment 1: Average evaluation scores for "dynamic"



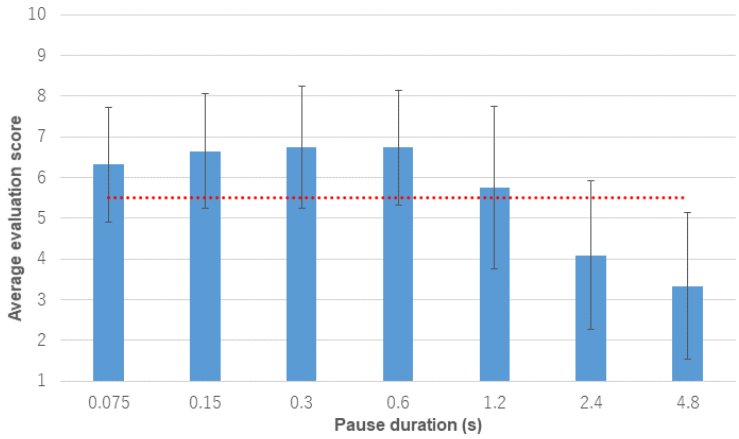
(g) Experiment 1: Average evaluation scores for "clear-cut" Japanese listeners (n=19)



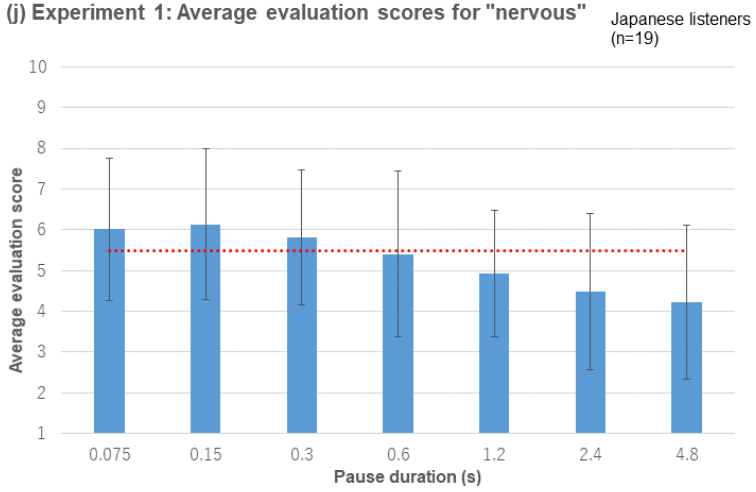
(h) Experiment 1: Average evaluation scores for "elegant" Japanese listeners (n=19)



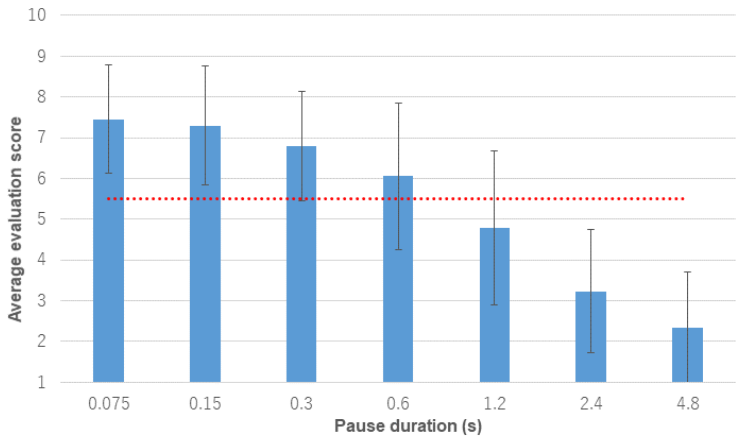
(i) Experiment 1: Average evaluation scores for "smooth" Japanese listeners (n=19)



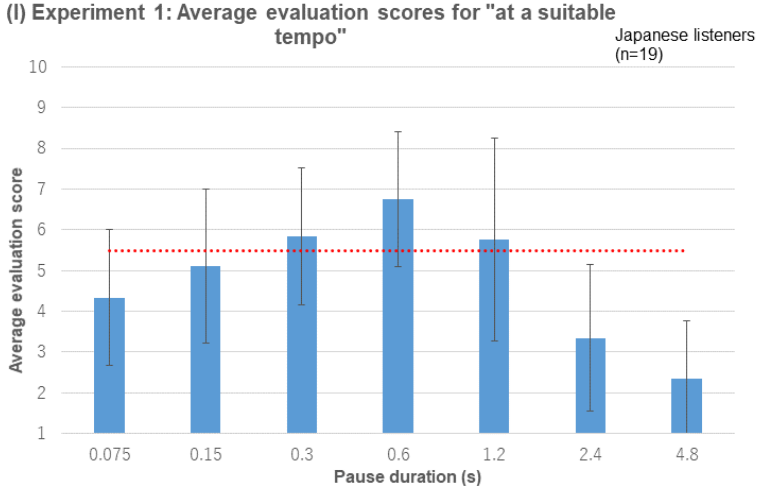
(j) Experiment 1: Average evaluation scores for "nervous"



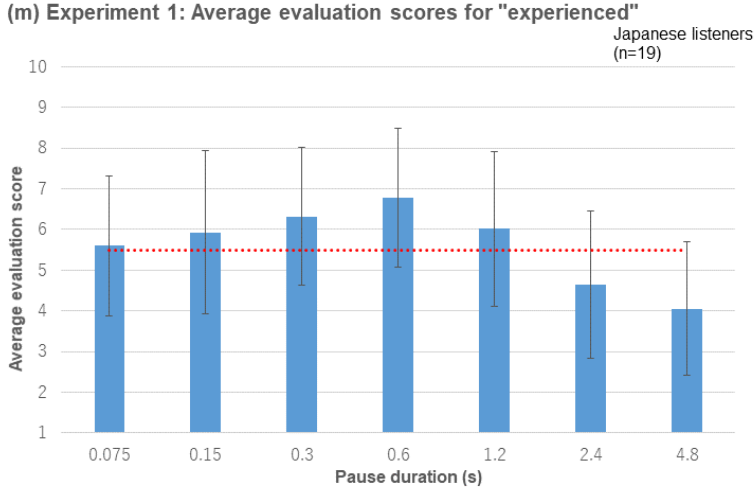
(k) Experiment 1: Average evaluation scores for "speedy"



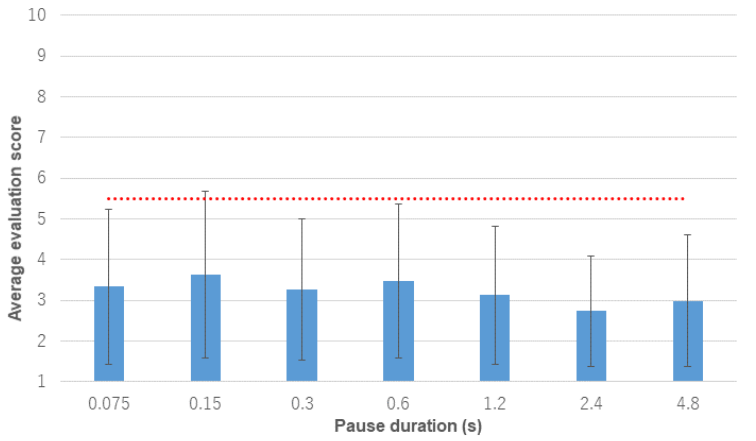
(l) Experiment 1: Average evaluation scores for "at a suitable tempo"



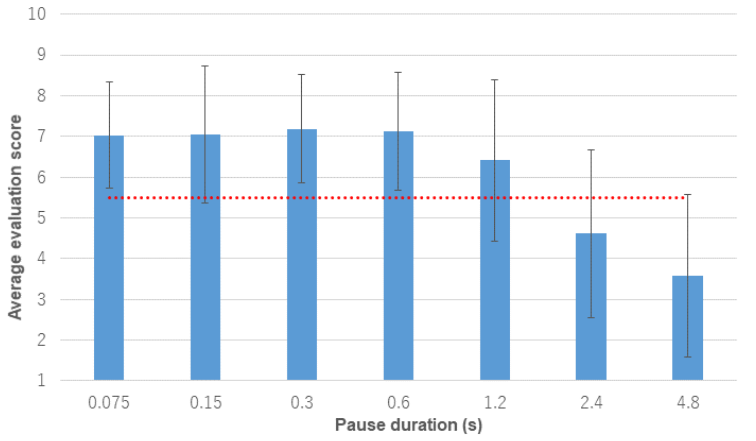
(m) Experiment 1: Average evaluation scores for "experienced"



(n) Experiment 1: Average evaluation scores for "shrill"

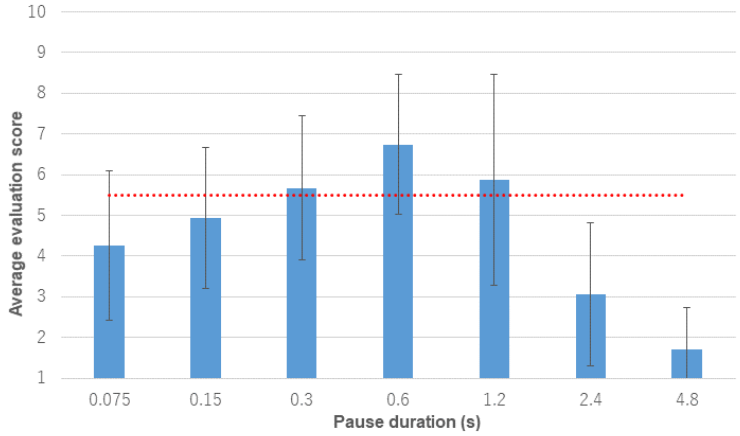


(o) Experiment 1: Average evaluation scores for "fluent"



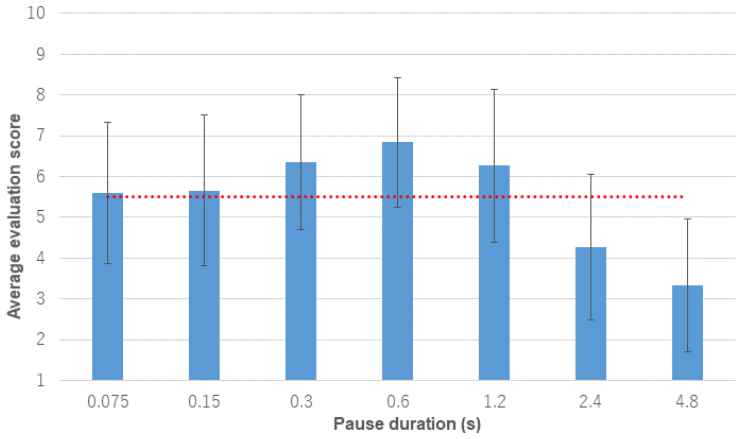
(p) Experiment 1: Average evaluation scores for "with appropriate pause duration"

Japanese listeners (n=19)



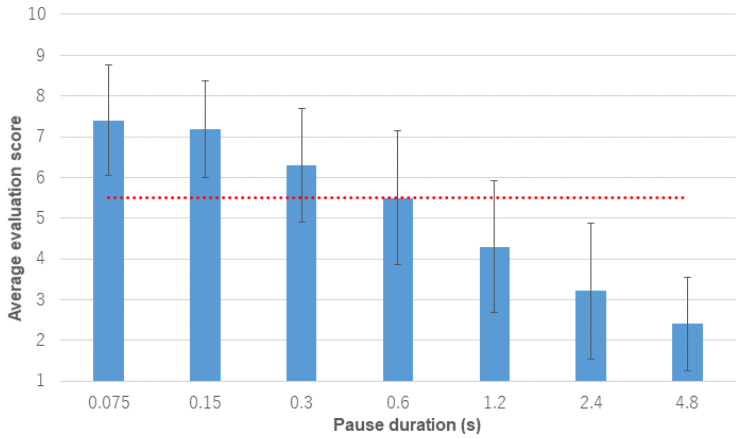
(q) Experiment 1: Average evaluation scores for "well-practiced"

Japanese listeners (n=19)

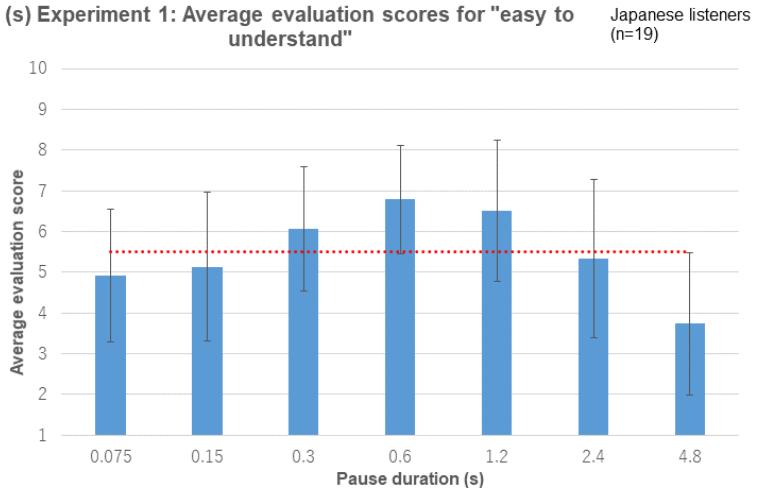


(r) Experiment 1: Average evaluation scores for "fast"

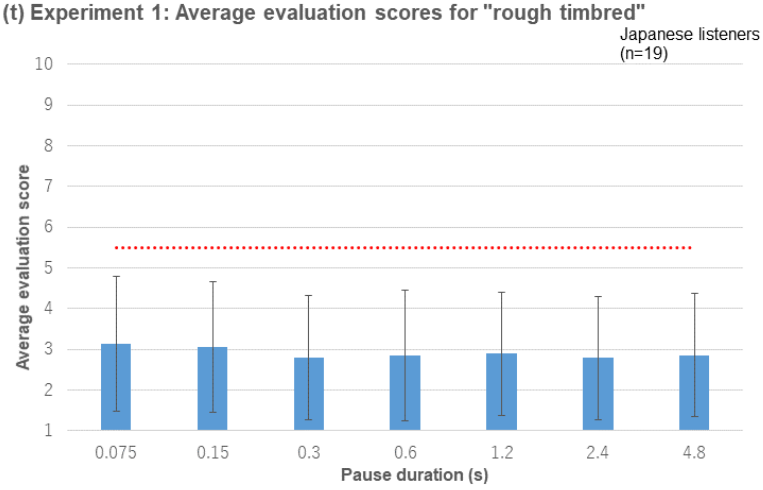
Japanese listeners (n=19)



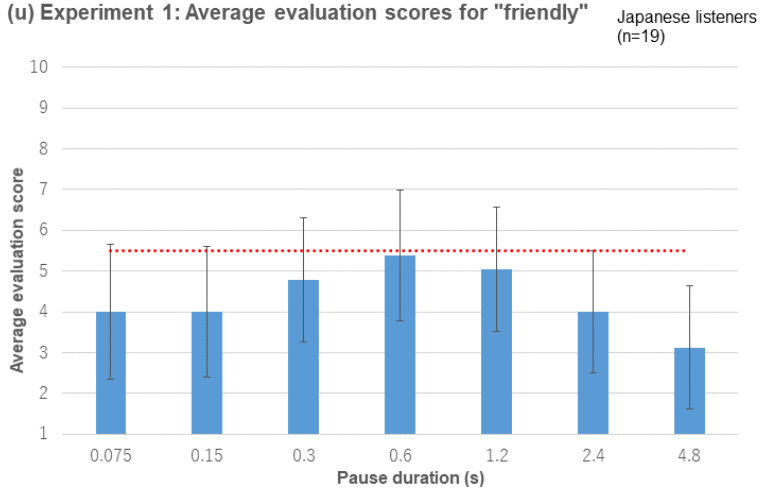
(s) Experiment 1: Average evaluation scores for "easy to understand"



(t) Experiment 1: Average evaluation scores for "rough timbred"

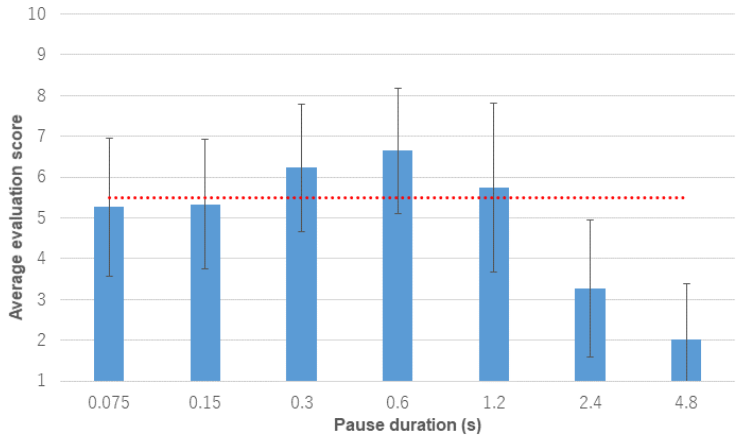


(u) Experiment 1: Average evaluation scores for "friendly"



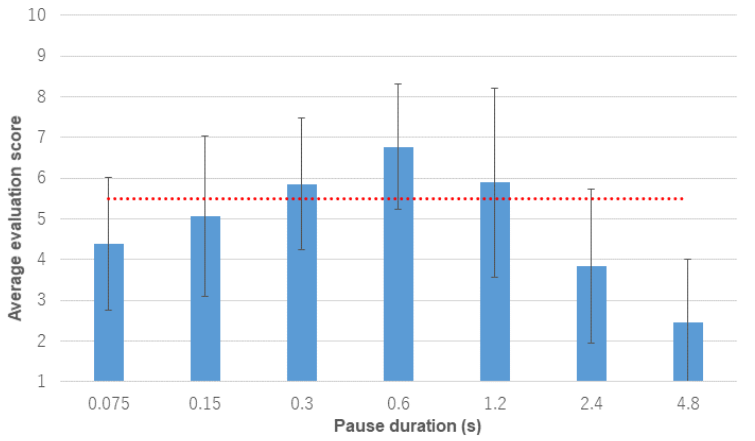
(v) Experiment 1: Average evaluation scores for "natural"

Japanese listeners (n=19)



(w) Experiment 1: Average evaluation scores for "with appropriate ryhtem"

Japanese listeners (n=19)



Appendix F. Informed consent and instructions of Experiment 2

KYUSHU UNIVERSITY

Human Science, Faculty of Design
4-9-1, Shiobaru, Minami-ku, Fukuoka
Japan
Prof Dr. Yoshitaka Nakajima

Consent Form

I, ...……......, give consent to my participation in the research
Name (please print)

project entitled **"Human perception and cognition: Analysis of speech from the viewpoint of perceptual psychology"**

In giving my consent, I acknowledge that:

1. The procedures required for the project and the time involved have been explained to me, and any questions I have about the project have been answered to my satisfaction.

2. I have read the Participant Information Statement and have been given the opportunity to discuss the information and involvement in the project with the researcher/s.

3. I understand that participation is voluntary and that I can withdraw from the study at any time, without penalty or prejudice.

4. I understand that my involvement is strictly confidential and no information about me and my data will be used in any way that reveals my identity, except when agreed otherwise.

…… Date …….
Full name (please print)

If you require further information, please contact :

Yoshitaka Nakajima
Telephone: +81 92 553 4564 / Fax: +81 92 553 4520
Address: 4-9-1, Shiobaru, Minami-ku, Fukuoka, 815-8540, Japan

Email: nakajima@design.kyushu-u.ac.jp

Instructions

Thank you for joining the experiment today.
From now on, English speeches will be presented.

How to operate the computer program.

3. After entering the lab, and the screen will be shown as below:

2. Press the "PLAY" button and the speech will be presented.
3. Please evaluate the evaluation items according to the answer sheet prepared on the table, after hearing the speech. For the evaluation method, please refer to the following "solution paper evaluation method".
4. Please press the "NEXT" button to enter the next stimulus screen when you finish evaluate the speech.
5. Please press the "PLAY" button to display the next stimulus.
6. It means that the session is finished, when the message "Thanks for your cooperation. Please take a rest!" presented on the screen. You can take a rest if you want or go on to the next session.

Evaluation method of the answer sheet.

1. There will be 12 evaluation items for one stimulus. Please judge how it is related to the speech. An example of evaluation item is shown as below.

Stimulus 1:

	not									very much
smooth	1	2	3	4	5	6	7	8	9	10

4. There will be 10 steps in each evaluation items. “1” presents for “not”, and “10” presents for “very much”.

3. When listen to the speeches, please imagine that you are an inspector of English public speaking. After listening to the speech, please judge the content of each evaluation project based on performance (speech style) of the speech.

4. Please turn off the answer sheet after finishing judging one stimulus. Please do not refer to previous evaluation scores.

5. Please circle the score on the answer sheet. Do not circle the space between the numbers, please select a number.

6. Before the experiment, please have a look at the 12 evaluation items.

Attentions

·After pressing the play button, Please do not touch the mouse until a series of speeches stop.

· Please don't look at the watch or any other things that can show the time.

·Please don't be hesitate to tell me if you do not feel very well or make a operation mistake. The experiment will not give you any harm, even if you do not finish the whole experiment.

· Please do not use your hands or any other part of your body to take a rhythm.

· There will be a practice session, session 1 and session 2 in this experiment. You can take a break between each section.

• The next stimuli will not be presented if you just press the "NEXT" button. Please make sure to press the “PLAY” button in order to present the next stimuli.

Thanks for your cooperation.

教示

本日は実験に参加していただき、ありがとうございます。
これから、英語スピーチ(演説、口頭発表、テーブルスピーチ等を含む)を呈示します。

パソコンを操作する手順

1. 実験室に入ると、パソコンのスクリーン画面で次のように表示されます。

2. 「PLAY」ボタンを押すと、音声が呈示されます。
3. 音声を聞いた後、机の上に用意してある回答用紙の評定項目に即して評定してください。評定する方法については以下の「回答用紙の評定方法」を参照してください。
4. 評定した後、「NEXT」ボタンを押すと、次の刺激の画面に入ります。
5. 「PLAY」ボタンを押すと、次の音声が呈示されます。
6. 画面に「お疲れ様でした。」というメッセージが表示されたら、セッションが終わるという意味です。休憩を取っても、次のセッションに進んでも結構です。

回答用紙を用いた評定方法

2. 一つの刺激に対して、12 項目の評定を行います。それぞれの項目がどのように音声に関わるかを判定してください。評定項目の例を以下に示します。

刺激 1：

全く当てはまらない　　非常によく当てはまる

滑らかな：　　1　2　3　4　5　6　7　8　9　10

3. 各評定項目の内容について“全く当てはまらない”(1)から“ 非常によく当てはまる”(10)まで、10 段階 （1－10）を設けています。
4. スピーチを聞く際には、自分が英語演説の指導員であると想像してください。スピーチを聞き終わってから，各評定項目の内容に即して話しかたの評定を行ってください。
5. 一つの刺激を評定し終わったら、回答用紙を裏返してください。前の評定点数を参照しないようにしてください。
6. 評定用紙に書かれた点数に○を付けてください。数字と数字の間には○を付けずに、必ず、いずれかの数字に○を付けてください。
7. 実験が始まる前に、12 の評価項目に目を通しておいてください。

注意事項

- 「PLAY」ボタンを押した後は、一連のスピーチの再生が終わるまでマウスに触らないでください。
- 時計など、時間のわかるものは見ないでください。
- 気分が優れなかったり、操作を間違えたりした場合は、遠慮なく実験者に申し出てください。実験を中断しても、実験自体に不都合が生じることはありません。
- 手足など、体の一部でリズムを取らないようにしてください。
- 本実験は、練習セッション、第 1 セッション、第 2 セッションの 3 つで構成されています。各セッションの間に休憩をとることができます。
- 「NEXT」ボタンを押しただけでは、次の音声は自動的には流れません。次の音声を流すためには、必ず「PLAY」ボタンを押してください。

それでは、よろしくお願いいたします。

实验指示

感谢您百忙之中抽出时间参加今天的实验。
从现在开始，将播放英语演讲包含演说，口头发表，席间致辞等。

如何操作计算机程序。

1.进入实验室后，计算机屏幕画面如下图所示:

2. 按“PLAY”键，播放音频。

3.音频播放完后，请根据桌上准备的答题纸上的评价项目对音频进行评分。评分方法请参考以下“答题纸评分方法”。

4. 当您完成评分时，请按下“NEXT”键进入下一个音频画面。

5. 请按“播放”键，以播放下一个音频。

8. 当屏幕提示“谢谢您的合作”时，表示此环节结束。你可以适当休息一下，或者继续下一环节。

答题纸评分方法。

1. 一个音频对应有 12 个评价项目。请对每项评价项目与所播放音频的关系进行评分。评价项目的示例如下所示。

音频 1:

	非常不符合									非常符合
流畅的	1	2	3	4	5	6	7	8	9	10

2.每个评价项目将有 10 个阶段。“1”表示“非常不符合不”，“10”表示“非常符合”。

3.听演讲音频时候，请想象你是一名英语演讲的指导老师。听完演讲音频后，请根据每项评价项目的内容对演讲的整体表现进行评分。

4. 完成一个音频的评分后，请反扣答题纸。评价当前音频时请不要参考前一个音频的评分内容。

5. 请圈答题纸上的显示的分数。不要圈数字之间的空格，请选择一个数字画圈。

6. 在实验之前，请浏览一下 12 个评价项目。

注意事项

•按下播放按钮后，请不要触摸鼠标，直到音频播放结束。

•请不要看手表或任何其他能显示时间的物品。

•如果你觉得不舒服或操作失误，不要有任何犹豫请告知主试。即使你没有完成整个实验，这个实验也不会给你带来任何伤害。

•请不要用你的手或身体的任何其他部位来敲击节奏。

•本实验将有一个练习环节，第 1 环节和第 2 环节。你可以在每个环节之间休息一下。

•如果您只按下“NEXT”按钮，音频不会自动播放。请务必按下“PLAY”按钮，以播放音频。

谢谢您的合作。

Stimuli	not									very much
with appropriate rhythm	1	2	3	4	5	6	7	8	9	10
rushed	1	2	3	4	5	6	7	8	9	10
natural	1	2	3	4	5	6	7	8	9	10
rough-timbred	1	2	3	4	5	6	7	8	9	10
skillful	1	2	3	4	5	6	7	8	9	10
speedy	1	2	3	4	5	6	7	8	9	10
at a suitable tempo	1	2	3	4	5	6	7	8	9	10
well-practiced	1	2	3	4	5	6	7	8	9	10
fast	1	2	3	4	5	6	7	8	9	10
with appropriate pause duration	1	2	3	4	5	6	7	8	9	10
friendly	1	2	3	4	5	6	7	8	9	10
high-pitched	1	2	3	4	5	6	7	8	9	10

刺激	非常不符合									非常符合
匆忙的	1	2	3	4	5	6	7	8	9	10
自然的	1	2	3	4	5	6	7	8	9	10
音质粗糙的	1	2	3	4	5	6	7	8	9	10
擅长的	1	2	3	4	5	6	7	8	9	10
有速度感的	1	2	3	4	5	6	7	8	9	10
节拍适当的	1	2	3	4	5	6	7	8	9	10
充分练习的	1	2	3	4	5	6	7	8	9	10
语速快的	1	2	3	4	5	6	7	8	9	10
停顿时长适当的	1	2	3	4	5	6	7	8	9	10
友好的	1	2	3	4	5	6	7	8	9	10
声调高的	1	2	3	4	5	6	7	8	9	10
节奏适当的	1	2	3	4	5	6	7	8	9	10

刺激	全くあてはまらない									非常によくあてはまる
リズムが適切である	1	2	3	4	5	6	7	8	9	10
せきたてるような	1	2	3	4	5	6	7	8	9	10
自然な	1	2	3	4	5	6	7	8	9	10
音質が荒い	1	2	3	4	5	6	7	8	9	10
上手な	1	2	3	4	5	6	7	8	9	10
スピード感がある	1	2	3	4	5	6	7	8	9	10
テンポが適切である	1	2	3	4	5	6	7	8	9	10
よく練習できている	1	2	3	4	5	6	7	8	9	10
速い	1	2	3	4	5	6	7	8	9	10
ポーズの長さが適切である	1	2	3	4	5	6	7	8	9	10
親しみやすい	1	2	3	4	5	6	7	8	9	10
高音である	1	2	3	4	5	6	7	8	9	10

问卷调查

姓名　　性别　　年龄

大学　　专业　　年级

何时开始学习英语， 日语？

分别学习了多少年？

是否有英语， 日语测验成绩， 分别多少分？ TOEIC， IELTS， TOFEL，N2

如何学习的日语？

在大学时是否有英语课，如果有英语课是什么样的英语课？

在日常生活中是否有机会使用英语，日语？如果有是什么样的场合？

对于自己现阶段的英语水平，有什么在意的点么？想要提高哪方面的能力？

是否受过英语公共演讲相关训练？如果有，是什么样的训练？

对于今天的实验，有何感想？

给自己的英语， 日语水平打分

英语　优秀（上　中 下）　良好(上　中　下）　及格（上　中　下）　不及格(上 中 下

日语　优秀（上　中 下）　良好(上　中　下）　及格（上　中　下）　不及格(上 中 下アンケート

名前：

性別：　　男性　　女性

年齢：

大学：　　専攻：　　学年：

いつから英語を勉強し始めましたか？

英語を何年ぐらい勉強していますか？

英語テストの成績ありますか？点数はいくらですか（およその値で結構です）？

（TOEIC，IELTS，TOEFL，GMAT など）

大学などで、英語の授業を受けていますか？もし受けている場合、どんな授業ですか？

普段の生活において、英語を使う機会ありますか？どんな場合に英語を話しますか？

自分の英語について、どんなことが気になっていますか？改善したいことはありますか？

（例えば：話す力、書く能力、読解力、文法、発音など）

パブリック・スピーキングに関する訓練を受けたことありますか？もし受けた場合、どんな訓練でしたか？

本日の評価用紙の評価項目について、感想などはありますか？

自分の英語力を判定してください。

　優(上・中・下)　良(上・中・下)　可(上・中・下)　不可(上・中・下

Appendix G. Sound pressure level (SPL) of stimuli in Experiment 2 (Fast peak)

Stimuli Number	1st (dB LAF)	2nd (dB LAF)	3rd (dB LAF)	Avg (dB LAF)
Pure Tone	62.0	62.0	62.0	62.0
0	68.7	68.7	68.7	68.7
1	68.7	68.7	68.7	68.7
2	68.7	68.7	68.7	68.7
3	68.7	68.7	68.7	68.7
4	68.7	68.7	68.7	68.7
5	68.7	68.7	68.7	68.7
6	68.7	68.7	68.7	68.7
7	68.7	68.7	68.7	68.7
8	68.7	68.7	68.7	68.7
9	68.7	68.7	68.7	68.7
10	68.7	68.7	68.7	68.7
11	68.7	68.7	68.7	68.7
12	68.7	68.7	68.7	68.7
13	68.7	68.7	68.7	68.7
14	68.7	68.7	68.7	68.7
15	68.7	68.7	68.7	68.7
16	68.7	68.7	68.7	68.7
17	68.7	68.7	68.7	68.7
18	68.7	68.7	68.7	68.7
19	68.7	68.7	68.7	68.7
20	68.7	68.7	68.7	68.7
21	68.7	68.7	68.7	68.7
22	68.7	68.7	68.7	68.7
23	68.7	68.7	68.7	68.7
24	68.7	68.7	68.7	68.7
25	68.7	68.7	68.7	68.7
26	68.7	68.7	68.7	68.7
27	66.2	66.2	66.2	66.2
28	66.2	66.2	66.2	66.2
29	66.2	66.2	66.2	66.2
30	66.2	66.2	66.2	66.2
31	66.2	66.2	66.2	66.2
32	66.2	66.2	66.2	66.2
33	66.2	66.2	66.2	66.2
34	66.2	66.2	66.2	66.2
35	66.2	66.2	66.2	66.2
36	66.2	66.2	66.2	66.2
37	66.2	66.2	66.2	66.2
38	66.2	66.2	66.2	66.2
39	66.2	66.2	66.2	66.2
40	66.2	66.2	66.2	66.2
41	66.2	66.2	66.2	66.2
42	66.2	66.2	66.2	66.2
43	66.2	66.2	66.2	66.2
44	66.2	66.2	66.2	66.2
45	66.2	66.2	66.2	66.2
46	66.2	66.2	66.2	66.2
47	66.2	66.2	66.2	66.2
48	66.2	66.2	66.2	66.2
49	66.2	66.2	66.2	66.2
50	66.2	66.2	66.2	66.2
51	66.2	66.2	66.2	66.2
52	66.2	66.2	66.2	66.2
53	66.2	66.2	66.2	66.2

Appendix H. Factor analysis of Experiment 2 data

English group

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.852
Bartlett's Test of Sphericity	Approx. Chi-Square	9833.681
	df	66
	Sig.	0.000

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared			Rotation Sums of Squared		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.789	39.912	39.912	4.789	39.912	39.912	4.750	39.581	39.581
2	2.937	24.473	64.385	2.937	24.473	64.385	2.650	22.081	61.663
3	1.083	9.028	73.413	1.083	9.028	73.413	1.410	11.750	73.413
4	0.899	7.492	80.905						
5	0.669	5.579	86.484						
6	0.488	4.067	90.551						
7	0.264	2.197	92.749						
8	0.235	1.955	94.704						
9	0.196	1.633	96.337						
10	0.169	1.405	97.742						
11	0.157	1.306	99.048						
12	0.114	0.952	100.000						

Extraction Method: Principal Component Analysis.

Rotated Component Matrix[a]

	Component		
	1	2	3
with appropriate rhythm	0.900	0.022	0.055
at a suitable tempo	0.892	-0.005	0.026
natural	0.870	0.062	0.134
with appropriate pause duration	0.864	-0.023	0.066
skillful	0.818	0.056	-0.162
well-practiced	0.816	0.081	-0.162
friendly	0.520	-0.107	0.505
speedy	0.097	0.918	0.211
fast	0.112	0.915	0.220
rushed	-0.079	0.910	0.063
high-pitched	-0.077	0.295	0.714
rough-timbred	-0.046	0.172	0.686

Extraction Method: Principal Component

a. Rotation converged in 5 iterations.

Chinese group

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.877
Bartlett's Test of Sphericity	Approx. Chi-Square	7918.856
	df	66
	Sig.	0.000

Total Variance Explained

Componen	Initial Eigenvalues			Extraction Sums of Squared			Rotation Sums of Squared		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.133	42.776	42.776	5.133	42.776	42.776	4.887	40.725	40.725
2	2.282	19.020	61.797	2.282	19.020	61.797	2.415	20.125	60.850
3	1.109	9.244	71.040	1.109	9.244	71.040	1.223	10.190	71.040
4	0.906	7.551	78.591						
5	0.551	4.595	83.187						
6	0.499	4.156	87.342						
7	0.399	3.327	90.669						
8	0.315	2.627	93.296						
9	0.261	2.173	95.470						
10	0.197	1.645	97.114						
11	0.177	1.478	98.592						
12	0.169	1.408	100.000						

Extraction Method: Principal Component Analysis.

Rotated Component Matrix[a]

	Component		
	1	2	3
节奏适当的	0.906	0.009	-0.005
节拍适当的	0.895	0.059	-0.003
停顿时长适当的	0.886	0.054	0.004
自然的	0.850	-0.007	0.079
友好的	0.739	-0.190	0.067
充分练习的	0.722	0.343	0.279
擅长的	0.697	0.331	0.360
语速快的	-0.003	0.903	0.025
匆忙的	-0.180	0.868	-0.101
有速度感的	0.371	0.728	0.155
音质粗糙的	0.089	0.167	-0.857
声调高的	0.160	0.134	0.486

Extraction Method: Principal Component Analysis.

a. Rotation converged in 4 iterations.

Japanese group

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.914
Bartlett's Test of Sphericity	Approx. Chi-Square	11324.181
	df	66
	Sig.	0.000

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared			Rotation Sums of Squared		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.919	49.328	49.328	5.919	49.328	49.328	5.714	47.619	47.619
2	2.678	22.319	71.647	2.678	22.319	71.647	2.883	24.028	71.647
3	0.954	7.954	79.601						
4	0.803	6.693	86.294						
5	0.342	2.854	89.148						
6	0.269	2.243	91.390						
7	0.219	1.827	93.217						
8	0.200	1.668	94.885						
9	0.184	1.529	96.414						
10	0.156	1.300	97.714						
11	0.144	1.199	98.913						
12	0.130	1.087	100.000						

Extraction Method: Principal Component Analysis.

Rotated Component Matrix[a]

	Component	
	1	2
テンポが適切である	0.919	0.083
ポーズの長さが適切である	0.912	0.043
リズムが適切である	0.900	0.099
上手な	0.898	0.162
自然な	0.893	0.108
親しみやすい	0.870	-0.178
よく練習できている	0.835	0.219
速い	0.106	0.923
せきたてるような	-0.042	0.914
スピード感がある	0.319	0.868
高音である	0.064	0.499
音質が荒い	-0.226	0.239

Extraction Method: Principal Component

a. Rotation converged in 3 iterations.

Appendix I. Statistical analysis of Experiment 2 data

Speech rate factor

English group

Normality test of Speech Rate factor

Tests of Normality

Pairs		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
Comma-pause	Period-pause	Statistic	df	Sig.	Statistic	df	Sig.
0.00	0.00	0.171	21	0.113	0.890	21	0.023
0.15	0.15	0.141	21	.200*	0.945	21	0.267
0.15	0.30	0.092	21	.200*	0.986	21	0.986
0.15	0.60	0.105	21	.200*	0.983	21	0.966
0.15	1.20	0.099	21	.200*	0.979	21	0.913
0.15	2.40	0.139	21	.200*	0.934	21	0.164
0.30	0.15	0.130	21	.200*	0.941	21	0.231
0.30	0.30	0.121	21	.200*	0.988	21	0.995
0.30	0.60	0.095	21	.200*	0.981	21	0.942
0.30	1.20	0.138	21	.200*	0.958	21	0.484
0.30	2.40	0.147	21	.200*	0.893	21	0.025
0.60	0.15	0.095	21	.200*	0.968	21	0.699
0.60	0.30	0.106	21	.200*	0.970	21	0.734
0.60	0.60	0.133	21	.200*	0.961	21	0.541
0.60	1.20	0.169	21	0.120	0.929	21	0.134
0.60	2.40	0.116	21	.200*	0.950	21	0.346
1.20	0.15	0.199	21	0.029	0.904	21	0.041
1.20	0.30	0.189	21	0.047	0.933	21	0.161
1.20	0.60	0.135	21	.200*	0.944	21	0.266
1.20	1.20	0.172	21	0.104	0.918	21	0.079
1.20	2.40	0.204	21	0.023	0.916	21	0.071
2.40	0.15	0.193	21	0.039	0.894	21	0.027
2.40	0.30	0.114	21	.200*	0.966	21	0.649
2.40	0.60	0.173	21	0.103	0.901	21	0.036
2.40	1.20	0.174	21	0.095	0.913	21	0.064
2.40	2.40	0.147	21	.200*	0.953	21	0.382
original		0.132	21	.200*	0.950	21	0.334

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Friedman test between manipulated pause duration stimulus types

Test Statistics[a]

N	21
Chi-Square	281.583
df	26
Asymp. Sig.	0.000

a. Friedman Test

Holm-Bonferroni test of speech rate factor

	Pairs		P value	α Value	Result
	Comma-pause	Period-pause			
Original	0	0	0.000	0.0019231	significant
Original	0.15	0.15	0.000	0.002	significant
Original	0.15	0.30	0.000	0.0020833	significant
Original	0.30	0.15	0.000	0.0021739	significant
Original	0.60	0.15	0.000	0.0022727	significant
Original	0.60	0.30	0.000	0.002381	significant
Original	0.30	0.30	0.000	0.0025	significant
Original	1.20	0.15	0.000	0.0026316	significant
Original	0.15	0.60	0.001	0.0027778	significant
Original	0.30	0.60	0.001	0.0029412	significant
Original	1.20	0.60	0.001	0.003125	significant
Original	0.15	1.20	0.002	0.0033333	significant
Original	1.20	0.30	0.003	0.0035714	significant
Original	2.40	0.15	0.013	0.0038462	not significant
Original	0.30	1.20	0.016	0.0041667	not significant
Original	2.40	2.40	0.021	0.0045455	not significant
Original	0.60	0.60	0.039	0.005	not significant
Original	1.20	2.40	0.079	0.0055556	not significant
Original	0.60	2.40	0.159	0.00625	not significant
Original	2.40	0.30	0.205	0.0071429	not significant
Original	0.60	1.20	0.289	0.0083333	not significant
Original	2.40	0.60	0.434	0.01	not significant
Original	2.40	1.20	0.434	0.0125	not significant
Original	0.30	2.40	0.664	0.0166667	not significant
Original	0.15	2.40	0.903	0.025	not significant
Original	1.20	1.20	0.903	0.05	not significant

Chinese group

Normality test of Speech Rate factor

Tests of Normality

Pairs		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
Comma-pause	Period-pause	Statistic	df	Sig.	Statistic	df	Sig.
0.00	0.00	0.161	20	0.184	0.940	20	0.235
0.15	0.15	0.170	20	0.131	0.939	20	0.232
0.15	0.30	0.130	20	.200*	0.953	20	0.415
0.15	0.60	0.120	20	.200*	0.942	20	0.263
0.15	1.20	0.223	20	0.010	0.923	20	0.112
0.15	2.40	0.112	20	.200*	0.950	20	0.372
0.30	0.15	0.141	20	.200*	0.952	20	0.402
0.30	0.30	0.131	20	.200*	0.933	20	0.174
0.30	0.60	0.124	20	.200*	0.988	20	0.993
0.30	1.20	0.134	20	.200*	0.967	20	0.697
0.30	2.40	0.151	20	.200*	0.929	20	0.150
0.60	0.15	0.154	20	.200*	0.968	20	0.709
0.60	0.30	0.157	20	.200*	0.959	20	0.522
0.60	0.60	0.238	20	0.004	0.908	20	0.058
0.60	1.20	0.103	20	.200*	0.981	20	0.943
0.60	2.40	0.149	20	.200*	0.959	20	0.523
1.20	0.15	0.148	20	.200*	0.942	20	0.261
1.20	0.30	0.087	20	.200*	0.984	20	0.973
1.20	0.60	0.139	20	.200*	0.975	20	0.855
1.20	1.20	0.127	20	.200*	0.960	20	0.536
1.20	2.40	0.139	20	.200*	0.889	20	0.026
2.40	0.15	0.128	20	.200*	0.957	20	0.482
2.40	0.30	0.180	20	0.089	0.944	20	0.286
2.40	0.60	0.169	20	0.138	0.855	20	0.006
2.40	1.20	0.102	20	.200*	0.969	20	0.742
2.40	2.40	0.118	20	.200*	0.911	20	0.067
original		0.176	20	0.104	0.955	20	0.453

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Friedman test between manipulated pause duration stimulus types

Test Statistics[a]

N	20
Chi-Square	323.505
df	26
Asymp. Sig.	0.000

a. Friedman Test

Holm-Bonferroni test of speech rate factor

Pairs			P value	α Value	Result
	Comma-pause	Period-pause			
Original	0.30	0.30	0.000	0.002	significant
Original	2.40	2.40	0.000	0.0020833	significant
Original	0.15	0.15	0.000	0.0021739	significant
Original	0.60	0.15	0.000	0.0022727	significant
Original	0.30	0.15	0.000	0.002381	significant
Original	1.20	0.15	0.000	0.0025	significant
Original	0.60	0.30	0.001	0.0026316	significant
Original	0.60	0.60	0.001	0.0027778	significant
Original	0.15	0.30	0.001	0.0029412	significant
Original	0.30	2.40	0.001	0.003125	significant
Original	1.20	0.30	0.001	0.0033333	significant
Original	1.20	2.40	0.002	0.0035714	significant
Original	0.30	0.60	0.002	0.0038462	significant
Original	0.15	0.60	0.004	0.0041667	significant
Original	0.30	1.20	0.025	0.0045455	not significant
Original	2.40	0.60	0.025	0.005	not significant
Original	0.60	1.20	0.052	0.0055556	not significant
Original	0.15	1.20	0.073	0.00625	not significant
Original	0.60	2.40	0.086	0.0071429	not significant
Original	2.40	0.30	0.093	0.0083333	not significant
Original	2.40	1.20	0.117	0.01	not significant
Original	2.40	0.15	0.145	0.0125	not significant
Original	0.15	2.40	0.332	0.0166667	not significant
Original	1.20	0.60	0.852	0.025	not significant
Original	1.20	1.20	0.852	0.05	not significant

Japanese group

Normality test of Speech Rate factor

Tests of Normality

Pairs		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
Comma-pause	Period-pause	Statistic	df	Sig.	Statistic	df	Sig.
0	0	0.123	20	.200*	0.938	20	0.219
0.15	0.15	0.133	20	.200*	0.957	20	0.485
0.15	0.3	0.122	20	.200*	0.968	20	0.702
0.15	0.6	0.123	20	.200*	0.964	20	0.618
0.15	1.2	0.152	20	.200*	0.940	20	0.241
0.15	2.4	0.116	20	.200*	0.971	20	0.777
0.3	0.15	0.109	20	.200*	0.964	20	0.626
0.3	0.3	0.119	20	.200*	0.955	20	0.455
0.3	0.6	0.086	20	.200*	0.977	20	0.891
0.3	1.2	0.113	20	.200*	0.956	20	0.465
0.3	2.4	0.109	20	.200*	0.958	20	0.500
0.6	0.15	0.129	20	.200*	0.966	20	0.676
0.6	0.3	0.138	20	.200*	0.945	20	0.304
0.6	0.6	0.152	20	.200*	0.964	20	0.617
0.6	1.2	0.102	20	.200*	0.972	20	0.791
0.6	2.4	0.129	20	.200*	0.961	20	0.572
1.2	0.15	0.106	20	.200*	0.945	20	0.297
1.2	0.3	0.105	20	.200*	0.987	20	0.992
1.2	0.6	0.135	20	.200*	0.966	20	0.675
1.2	1.2	0.094	20	.200*	0.981	20	0.950
1.2	2.4	0.119	20	.200*	0.965	20	0.658
2.4	0.15	0.120	20	.200*	0.956	20	0.463
2.4	0.3	0.128	20	.200*	0.968	20	0.718
2.4	0.6	0.192	20	0.052	0.908	20	0.059
2.4	1.2	0.159	20	0.199	0.949	20	0.348
2.4	2.4	0.090	20	.200*	0.977	20	0.887
original		0.075	20	.200*	0.971	20	0.779

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Friedman test between manipulated pause duration stimulus types

Test Statistics[a]	
N	20
Chi-Square	395.541
df	26
Asymp. Sig.	0.000

a. Friedman Test

Holm-Bonferroni test of speech rate factor

	Pairs		P value	α Value	Result
	Comma-pause	Period-pause			
Original	0	0	0.000	0.0019231	significant
Original	0.15	0.15	0.000	0.002	significant
Original	0.15	0.30	0.000	0.0020833	significant
Original	0.15	0.60	0.000	0.0021739	significant
Original	0.30	0.15	0.000	0.0022727	significant
Original	0.30	0.30	0.000	0.002381	significant
Original	0.60	0.30	0.000	0.0025	significant
Original	0.60	0.15	0.000	0.0026316	significant
Original	0.30	0.60	0.000	0.0027778	significant
Original	1.20	0.30	0.000	0.0029412	significant
Original	1.20	0.15	0.000	0.003125	significant
Original	0.60	0.60	0.000	0.0033333	significant
Original	0.30	1.20	0.001	0.0035714	significant
Original	2.40	0.15	0.002	0.0038462	significant
Original	0.15	1.20	0.002	0.0041667	significant
Original	1.20	0.60	0.014	0.0045455	not significant
Original	1.20	1.20	0.017	0.005	not significant
Original	1.20	2.40	0.021	0.0055556	not significant
Original	0.60	1.20	0.028	0.00625	not significant
Original	2.40	2.40	0.040	0.0071429	not significant
Original	0.60	2.40	0.062	0.0083333	not significant
Original	2.40	0.30	0.093	0.01	not significant
Original	2.40	1.20	0.204	0.0125	not significant
Original	2.40	0.60	0.575	0.0166667	not significant
Original	0.30	2.40	0.654	0.025	not significant
Original	0.15	2.40	0.737	0.05	not significant

Naturalness factor

English group

Normality test of Naturalness factor

Tests of Normality

Pairs		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
Comma-pause	Period-pause	Statistic	df	Sig.	Statistic	df	Sig.
0.00	0.00	0.122	21	.200*	0.955	21	0.426
0.15	0.15	0.125	21	.200*	0.970	21	0.726
0.15	0.30	0.147	21	.200*	0.944	21	0.262
0.15	0.60	0.124	21	.200*	0.987	21	0.991
0.15	1.20	0.163	21	0.150	0.924	21	0.105
0.15	2.40	0.179	21	0.077	0.946	21	0.289
0.30	0.15	0.122	21	.200*	0.942	21	0.238
0.30	0.30	0.076	21	.200*	0.983	21	0.960
0.30	0.60	0.148	21	.200*	0.963	21	0.580
0.30	1.20	0.132	21	.200*	0.965	21	0.622
0.30	2.40	0.097	21	.200*	0.957	21	0.462
0.60	0.15	0.148	21	.200*	0.970	21	0.742
0.60	0.30	0.107	21	.200*	0.960	21	0.517
0.60	0.60	0.159	21	0.181	0.939	21	0.208
0.60	1.20	0.144	21	.200*	0.957	21	0.456
0.60	2.40	0.089	21	.200*	0.979	21	0.912
1.20	0.15	0.119	21	.200*	0.958	21	0.477
1.20	0.30	0.090	21	.200*	0.968	21	0.699
1.20	0.60	0.112	21	.200*	0.959	21	0.505
1.20	1.20	0.127	21	.200*	0.962	21	0.549
1.20	2.40	0.148	21	.200*	0.935	21	0.177
2.40	0.15	0.201	21	0.026	0.908	21	0.051
2.40	0.30	0.078	21	.200*	0.984	21	0.973
2.40	0.60	0.179	21	0.077	0.852	21	0.005
2.40	1.20	0.122	21	.200*	0.952	21	0.365
2.40	2.40	0.146	21	.200*	0.968	21	0.695
original		0.132	21	.200*	0.937	21	0.191

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Friedman test between manipulated pause duration stimulus types

Test Statistics[a]

N	21
Chi-Square	328.724
df	26
Asymp. Sig.	0.000

a. Friedman Test

Holm-Bonferroni test of Naturalness factor

	Pairs		P value	α Value	Result
	Comma-pause	Period-pause			
Original	0.60	2.40	0.000	0.0019231	significant
Original	1.20	0.15	0.000	0.002	significant
Original	2.40	0.15	0.000	0.0020833	significant
Original	2.40	0.60	0.000	0.0021739	significant
Original	2.40	1.20	0.000	0.0022727	significant
Original	2.40	2.40	0.000	0.002381	significant
Original	0.15	2.40	0.000	0.0025	significant
Original	0.30	2.40	0.000	0.0026316	significant
Original	2.40	0.30	0.000	0.0027778	significant
Original	1.20	2.40	0.000	0.0029412	significant
Original	0.00	0.00	0.000	0.003125	significant
Original	1.20	0.60	0.000	0.0033333	significant
Original	1.20	0.30	0.000	0.0035714	significant
Original	1.20	1.20	0.002	0.0038462	significant
Original	0.60	0.15	0.006	0.0041667	not significant
Original	0.15	0.15	0.011	0.0045455	not significant
Original	0.30	0.15	0.021	0.005	not significant
Original	0.30	0.30	0.023	0.0055556	not significant
Original	0.60	0.30	0.027	0.00625	not significant
Original	0.15	1.20	0.033	0.0071429	not significant
Original	0.15	0.30	0.050	0.0083333	not significant
Original	0.30	1.20	0.054	0.01	not significant
Original	0.60	1.20	0.159	0.0125	not significant
Original	0.15	0.60	0.566	0.0166667	not significant
Original	0.30	0.60	0.876	0.025	not significant
Original	0.60	0.60	0.876	0.05	not significant

Chinese group

Normality test of Naturalness factor

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
Comma-pause	Period-pause	Statistic	df	Sig.	Statistic	df	Sig.
0.00	0.00	0.153	20	.200*	0.921	20	0.103
0.15	0.15	0.121	20	.200*	0.965	20	0.639
0.15	0.30	0.142	20	.200*	0.965	20	0.648
0.15	0.60	0.187	20	0.066	0.909	20	0.060
0.15	1.20	0.133	20	.200*	0.964	20	0.617
0.15	2.40	0.120	20	.200*	0.966	20	0.666
0.30	0.15	0.084	20	.200*	0.984	20	0.971
0.30	0.30	0.187	20	0.064	0.950	20	0.368
0.30	0.60	0.106	20	.200*	0.967	20	0.687
0.30	1.20	0.105	20	.200*	0.967	20	0.697
0.30	2.40	0.131	20	.200*	0.967	20	0.687
0.60	0.15	0.131	20	.200*	0.961	20	0.554
0.60	0.30	0.112	20	.200*	0.981	20	0.945
0.60	0.60	0.173	20	0.119	0.958	20	0.507
0.60	1.20	0.111	20	.200*	0.970	20	0.765
0.60	2.40	0.123	20	.200*	0.970	20	0.764
1.20	0.15	0.130	20	.200*	0.965	20	0.654
1.20	0.30	0.129	20	.200*	0.949	20	0.354
1.20	0.60	0.160	20	0.190	0.939	20	0.234
1.20	1.20	0.123	20	.200*	0.959	20	0.528
1.20	2.40	0.109	20	.200*	0.975	20	0.846
2.40	0.15	0.108	20	.200*	0.983	20	0.970
2.40	0.30	0.131	20	.200*	0.941	20	0.247
2.40	0.60	0.160	20	0.191	0.958	20	0.498
2.40	1.20	0.099	20	.200*	0.985	20	0.984
2.40	2.40	0.233	20	0.006	0.912	20	0.069
original		0.106	20	.200*	0.966	20	0.675

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Friedman test between manipulated pause duration stimulus types

Test Statistics[a]	
N	20
Chi-Square	289.295
df	26
Asymp. Sig.	0.000

a. Friedman Test

Holm-Bonferroni test of Naturalness factor

	Pairs		P value	α Value	Result
	Comma-pause	Period-pause			
Original	0.00	0.00	0.000	0.0019231	significant
Original	0.15	2.40	0.000	0.002	significant
Original	0.30	2.40	0.000	0.0020833	significant
Original	1.20	2.40	0.000	0.0021739	significant
Original	2.40	0.15	0.000	0.0022727	significant
Original	2.40	0.30	0.000	0.002381	significant
Original	2.40	0.60	0.000	0.0025	significant
Original	2.40	2.40	0.000	0.0026316	significant
Original	0.60	2.40	0.000	0.0027778	significant
Original	1.20	0.15	0.000	0.0029412	significant
Original	2.40	1.20	0.000	0.003125	significant
Original	0.15	0.15	0.000	0.0033333	significant
Original	0.30	0.15	0.001	0.0035714	significant
Original	0.60	0.15	0.001	0.0038462	significant
Original	0.15	0.30	0.002	0.0041667	significant
Original	0.30	0.30	0.006	0.0045455	not significant
Original	1.20	0.30	0.010	0.005	not significant
Original	0.60	0.30	0.025	0.0055556	not significant
Original	0.30	0.60	0.057	0.00625	not significant
Original	0.15	0.60	0.093	0.0071429	not significant
Original	0.15	1.20	0.108	0.0083333	not significant
Original	1.20	0.60	0.117	0.01	not significant
Original	1.20	1.20	0.117	0.0125	not significant
Original	0.30	1.20	0.232	0.0166667	not significant
Original	0.60	0.60	0.411	0.025	not significant
Original	0.60	1.20	0.526	0.05	not significant

Japanese group

Normality test of Naturalness factor

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
Comma-pause	Period-pause	Statistic	df	Sig.	Statistic	df	Sig.
0	0	0.201	20	0.034	0.841	20	0.004
0.15	0.15	0.157	20	.200*	0.907	20	0.055
0.15	0.3	0.080	20	.200*	0.987	20	0.991
0.15	0.6	0.136	20	.200*	0.918	20	0.093
0.15	1.2	0.110	20	.200*	0.979	20	0.922
0.15	2.4	0.117	20	.200*	0.970	20	0.764
0.3	0.15	0.166	20	0.150	0.923	20	0.115
0.3	0.3	0.125	20	.200*	0.949	20	0.359
0.3	0.6	0.132	20	.200*	0.968	20	0.718
0.3	1.2	0.133	20	.200*	0.951	20	0.382
0.3	2.4	0.118	20	.200*	0.950	20	0.363
0.6	0.15	0.215	20	0.016	0.892	20	0.029
0.6	0.3	0.127	20	.200*	0.961	20	0.573
0.6	0.6	0.089	20	.200*	0.983	20	0.963
0.6	1.2	0.133	20	.200*	0.972	20	0.805
0.6	2.4	0.093	20	.200*	0.963	20	0.597
1.2	0.15	0.079	20	.200*	0.992	20	1.000
1.2	0.3	0.108	20	.200*	0.973	20	0.812
1.2	0.6	0.119	20	.200*	0.957	20	0.481
1.2	1.2	0.126	20	.200*	0.959	20	0.515
1.2	2.4	0.121	20	.200*	0.954	20	0.428
2.4	0.15	0.138	20	.200*	0.961	20	0.563
2.4	0.3	0.168	20	0.143	0.957	20	0.492
2.4	0.6	0.113	20	.200*	0.979	20	0.921
2.4	1.2	0.127	20	.200*	0.949	20	0.359
2.4	2.4	0.141	20	.200*	0.951	20	0.380
original		0.132	20	.200*	0.958	20	0.496

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Friedman test between manipulated pause duration stimulus types

Test Statistics[a]

N	20
Chi-Square	336.190
df	26
Asymp. Sig.	0.000

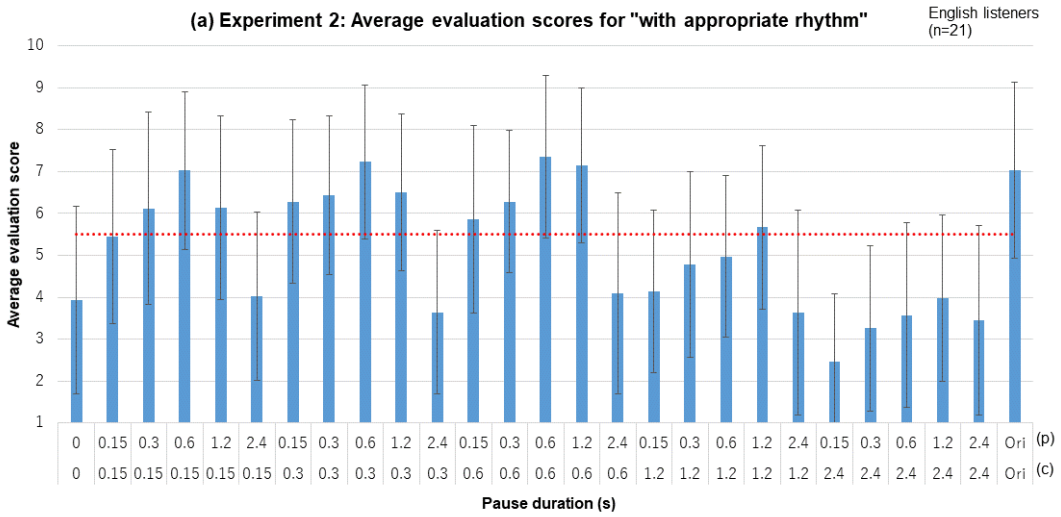
a. Friedman Test

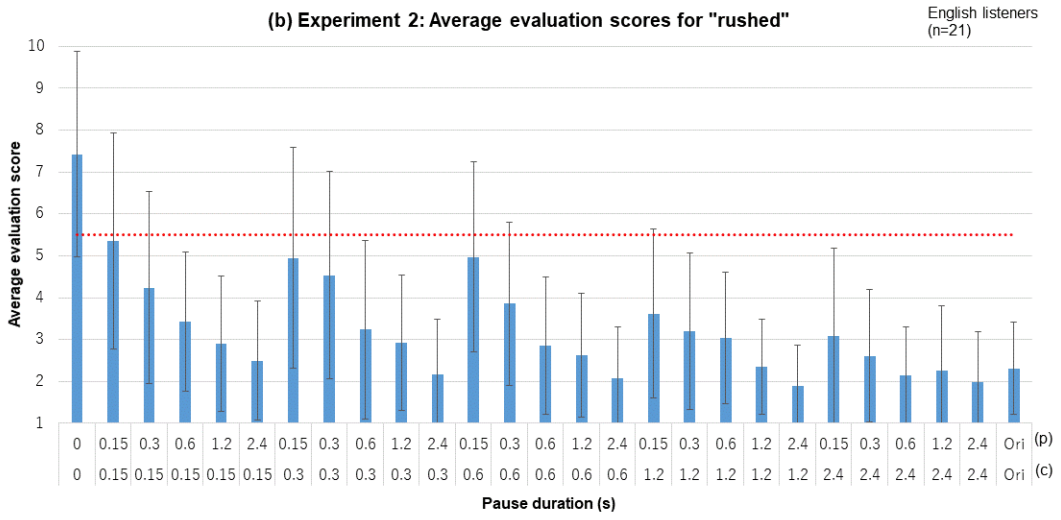
Holm-Bonferroni test of Naturalness factor

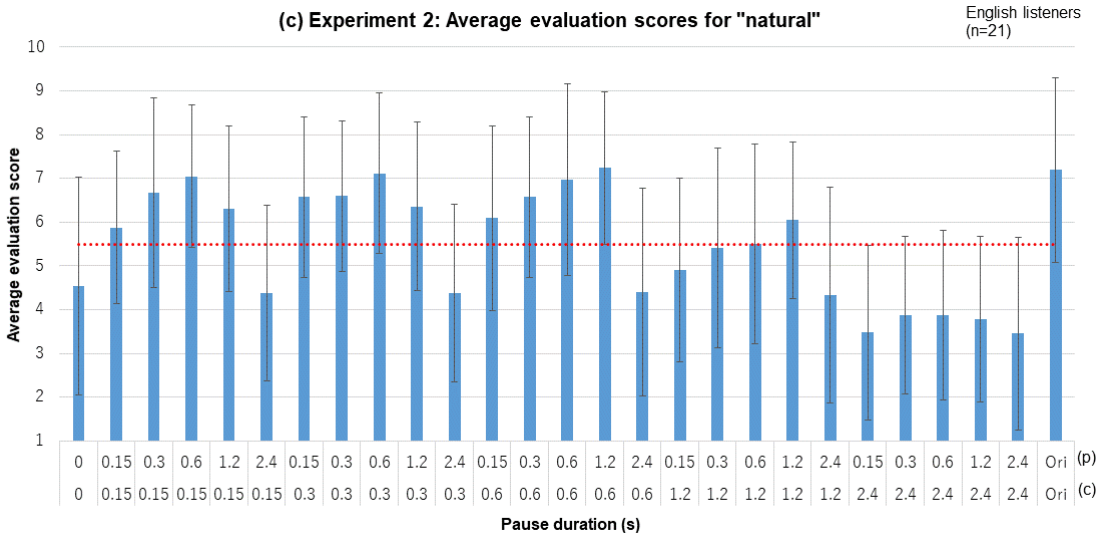
	Pairs		P value	α Value	Result
	Comma-pause	Period-pause			
Original	0.15	0.15	0.000	0.0019231	significant
Original	0.15	2.40	0.000	0.002	significant
Original	0.30	2.40	0.000	0.0020833	significant
Original	0.60	2.40	0.000	0.0021739	significant
Original	1.20	2.40	0.000	0.0022727	significant
Original	2.40	0.15	0.000	0.002381	significant
Original	2.40	0.30	0.000	0.0025	significant
Original	2.40	0.60	0.000	0.0026316	significant
Original	2.40	1.20	0.000	0.0027778	significant
Original	2.40	2.40	0.000	0.0029412	significant
Original	0	0	0.000	0.003125	significant
Original	1.20	0.15	0.000	0.0033333	significant
Original	0.60	0.15	0.000	0.0035714	significant
Original	1.20	0.30	0.001	0.0038462	significant
Original	0.30	0.15	0.002	0.0041667	significant
Original	0.15	0.30	0.002	0.0045455	significant
Original	0.30	0.30	0.010	0.005	not significant
Original	1.20	1.20	0.014	0.0055556	not significant
Original	0.30	1.20	0.044	0.00625	not significant
Original	0.60	0.60	0.057	0.0071429	not significant
Original	1.20	0.60	0.086	0.0083333	not significant
Original	0.15	1.20	0.145	0.01	not significant
Original	0.60	0.30	0.204	0.0125	not significant
Original	0.60	1.20	0.263	0.0166667	not significant
Original	0.30	0.60	0.709	0.025	not significant
Original	0.15	0.60	0.852	0.05	not significant

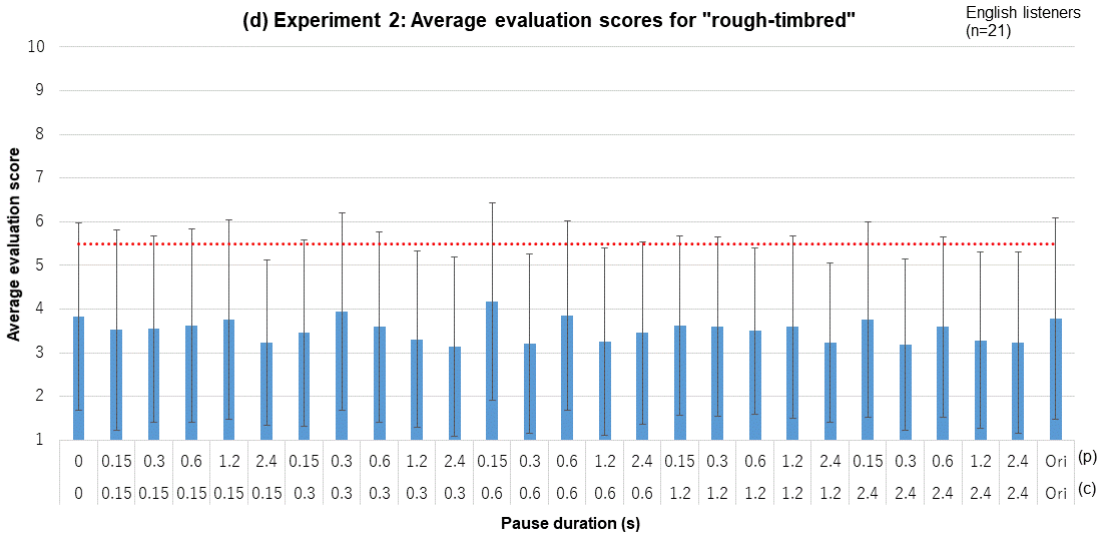
Appendix J. Average evaluation scores for each evaluation item of Experiment 2

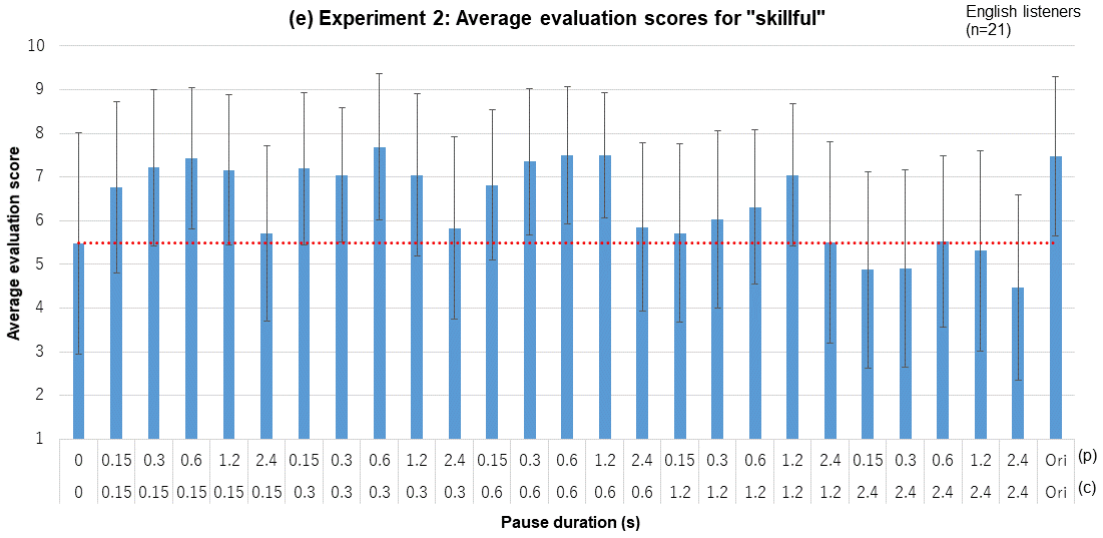
English group

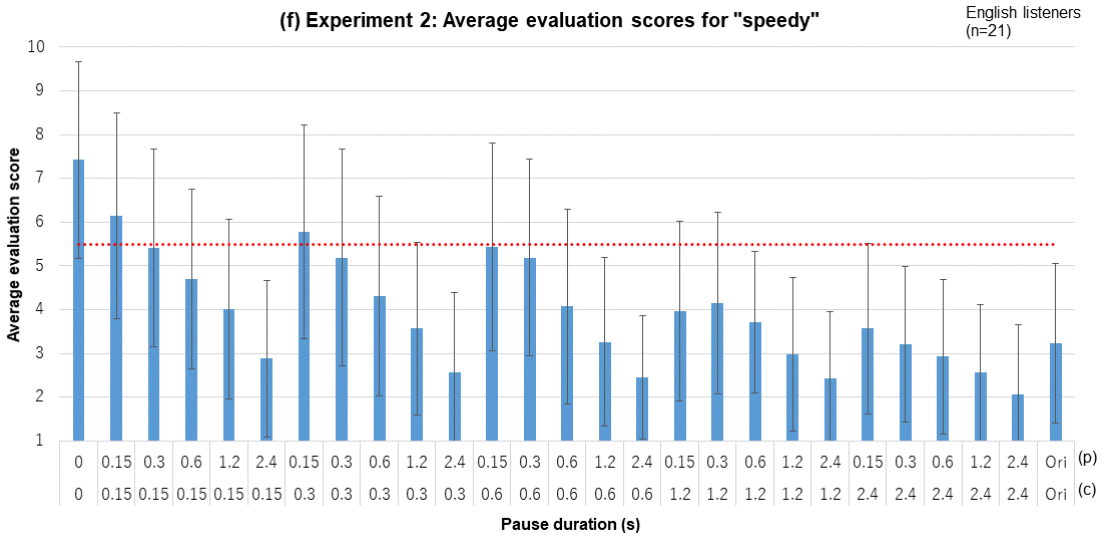


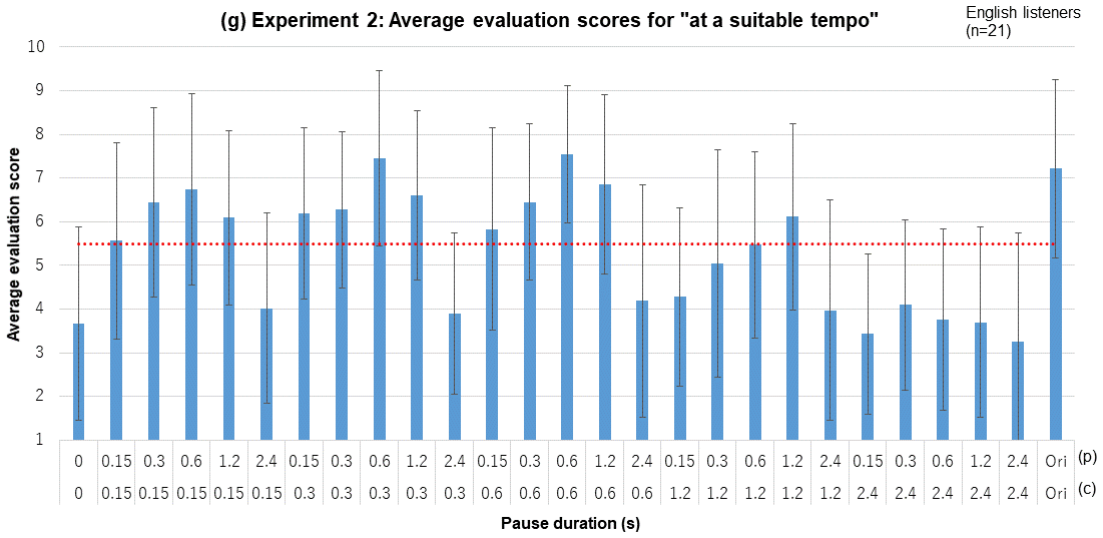


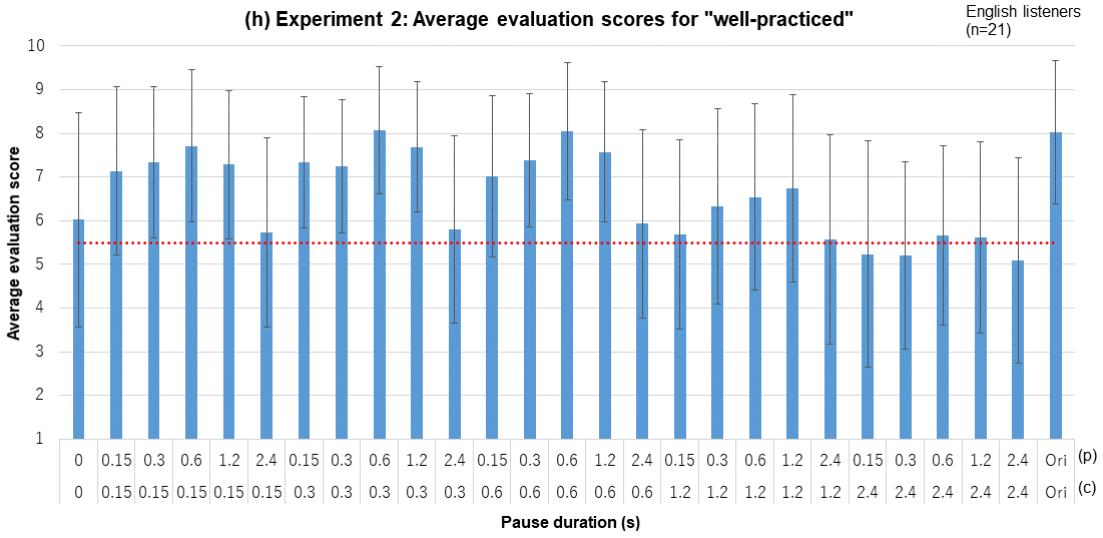

(c) Experiment 2: Average evaluation scores for "natural"
English listeners (n=21)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
0 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 Ori (p)
0 0.15 0.15 0.15 0.15 0.15 0.3 0.3 0.3 0.3 0.3 0.6 0.6 0.6 0.6 0.6 1.2 1.2 1.2 1.2 1.2 2.4 2.4 2.4 2.4 2.4 Ori (c)
Pause duration (s)

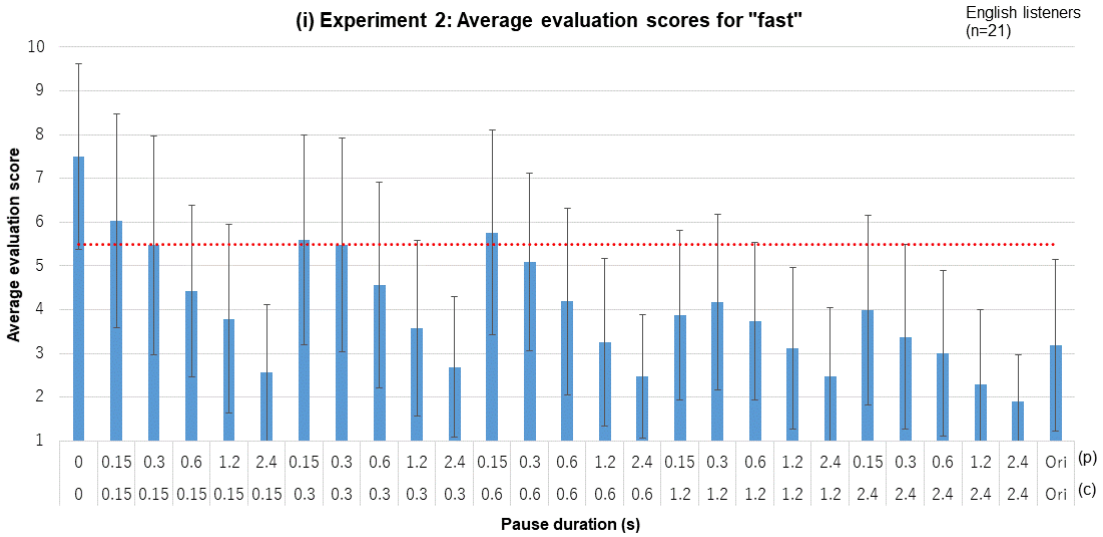

(d) Experiment 2: Average evaluation scores for "rough-timbred"
English listeners (n=21)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
0 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 Ori (p)
0 0.15 0.15 0.15 0.15 0.15 0.3 0.3 0.3 0.3 0.3 0.6 0.6 0.6 0.6 0.6 1.2 1.2 1.2 1.2 1.2 2.4 2.4 2.4 2.4 2.4 Ori (c)
Pause duration (s)

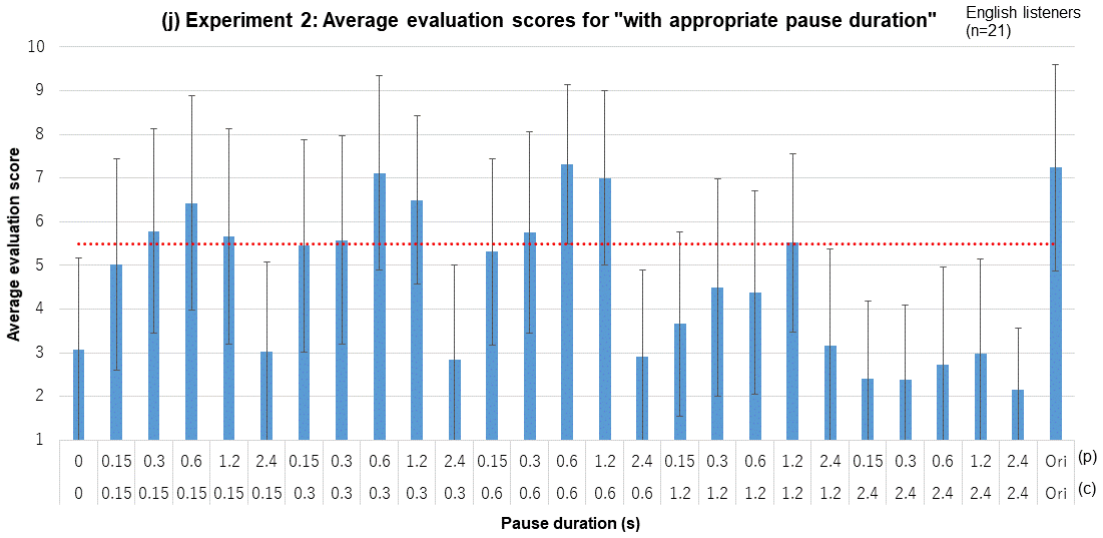

(e) Experiment 2: Average evaluation scores for "skillful"
English listeners (n=21)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
0 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 Ori (p)
0 0.15 0.15 0.15 0.15 0.15 0.3 0.3 0.3 0.3 0.3 0.6 0.6 0.6 0.6 0.6 1.2 1.2 1.2 1.2 1.2 2.4 2.4 2.4 2.4 2.4 Ori (c)
Pause duration (s)

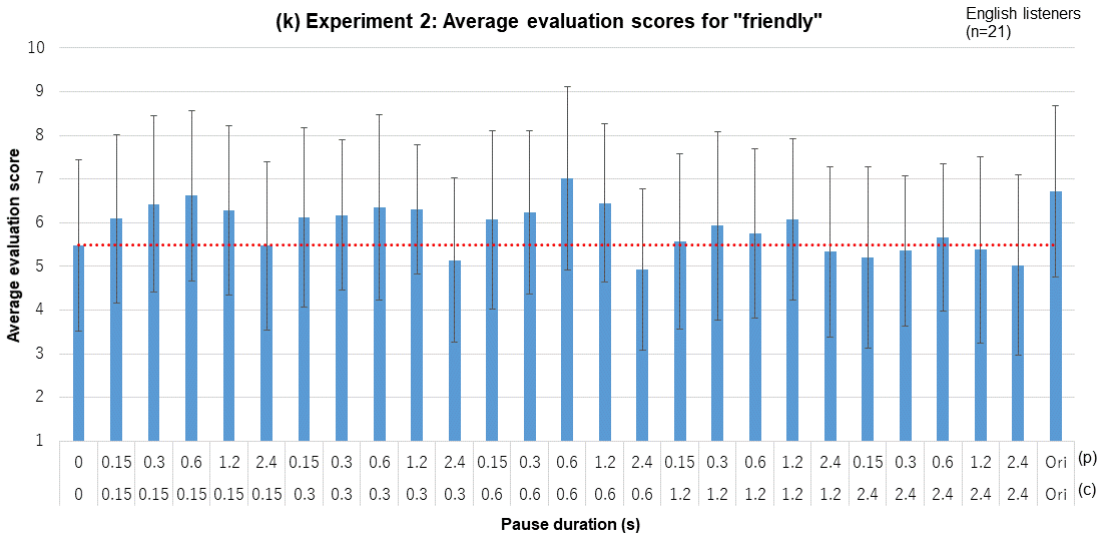

(f) Experiment 2: Average evaluation scores for "speedy"
English listeners (n=21)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
0 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 Ori (p)
0 0.15 0.15 0.15 0.15 0.15 0.3 0.3 0.3 0.3 0.3 0.6 0.6 0.6 0.6 0.6 1.2 1.2 1.2 1.2 1.2 2.4 2.4 2.4 2.4 2.4 Ori (c)
Pause duration (s)

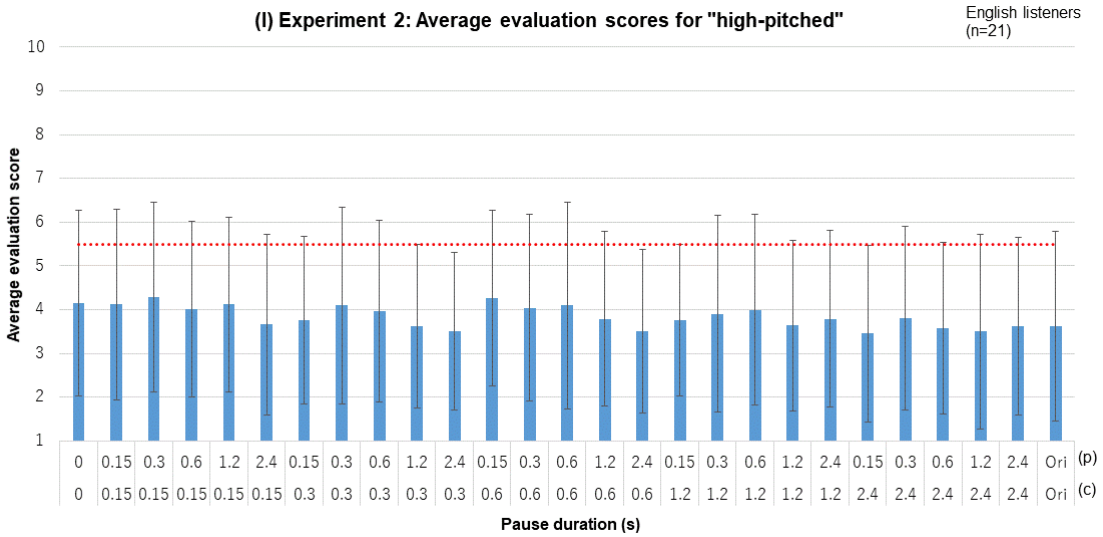

(g) Experiment 2: Average evaluation scores for "at a suitable tempo"
English listeners (n=21)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
0 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 Ori (p)
0 0.15 0.15 0.15 0.15 0.15 0.3 0.3 0.3 0.3 0.3 0.6 0.6 0.6 0.6 0.6 1.2 1.2 1.2 1.2 1.2 2.4 2.4 2.4 2.4 2.4 Ori (c)
Pause duration (s)


(h) Experiment 2: Average evaluation scores for "well-practiced"
English listeners (n=21)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
0 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 Ori (p)
0 0.15 0.15 0.15 0.15 0.15 0.3 0.3 0.3 0.3 0.3 0.6 0.6 0.6 0.6 0.6 1.2 1.2 1.2 1.2 1.2 2.4 2.4 2.4 2.4 2.4 Ori (c)
Pause duration (s)

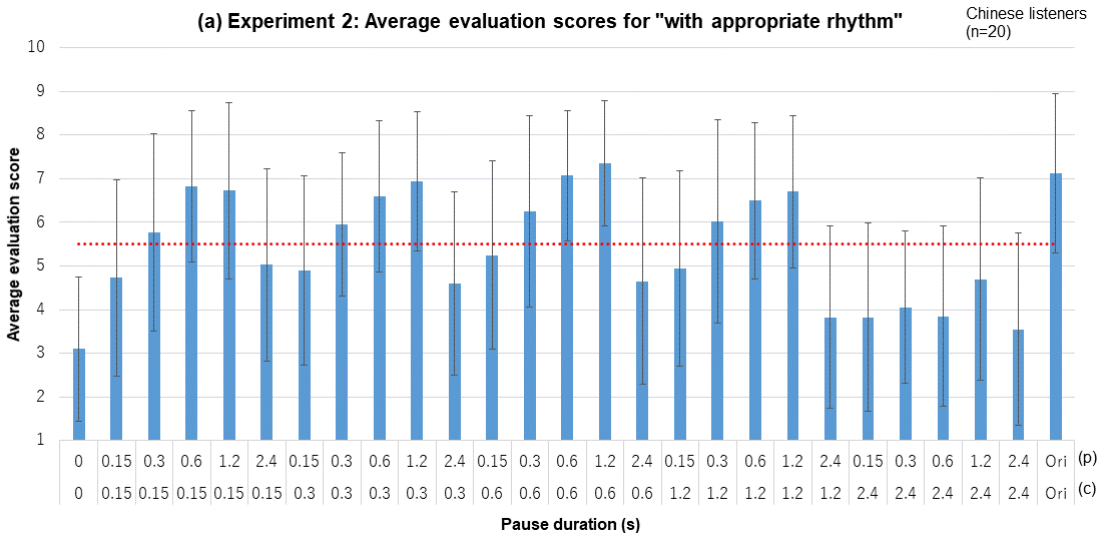

(i) Experiment 2: Average evaluation scores for "fast"
English listeners (n=21)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
0 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 Ori (p)
0 0.15 0.15 0.15 0.15 0.15 0.3 0.3 0.3 0.3 0.3 0.6 0.6 0.6 0.6 0.6 1.2 1.2 1.2 1.2 1.2 2.4 2.4 2.4 2.4 2.4 Ori (c)
Pause duration (s)

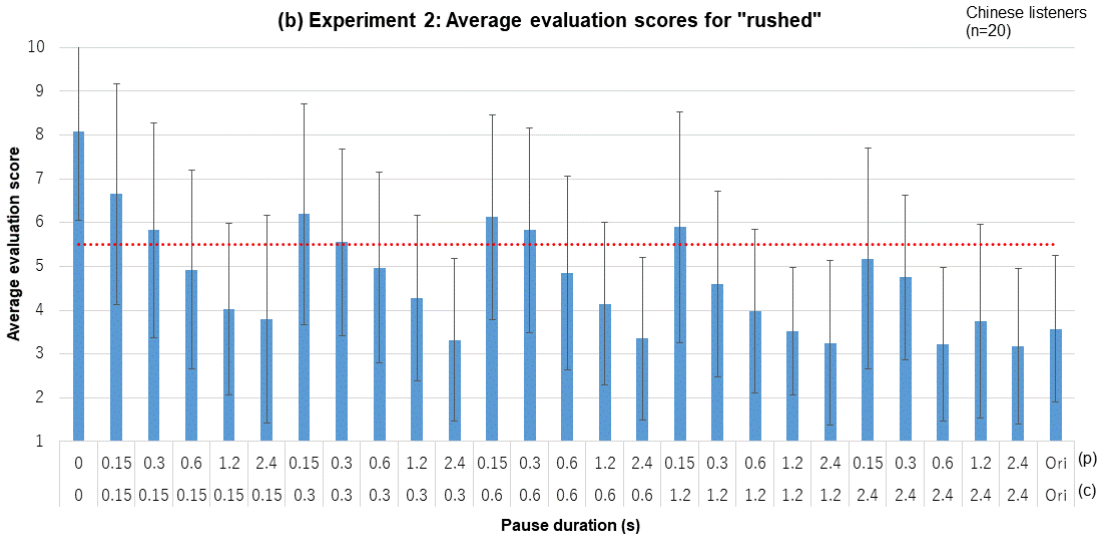

(j) Experiment 2: Average evaluation scores for "with appropriate pause duration"
English listeners (n=21)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
0 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 Ori (p)
0 0.15 0.15 0.15 0.15 0.15 0.3 0.3 0.3 0.3 0.3 0.6 0.6 0.6 0.6 0.6 1.2 1.2 1.2 1.2 1.2 2.4 2.4 2.4 2.4 2.4 Ori (c)
Pause duration (s)

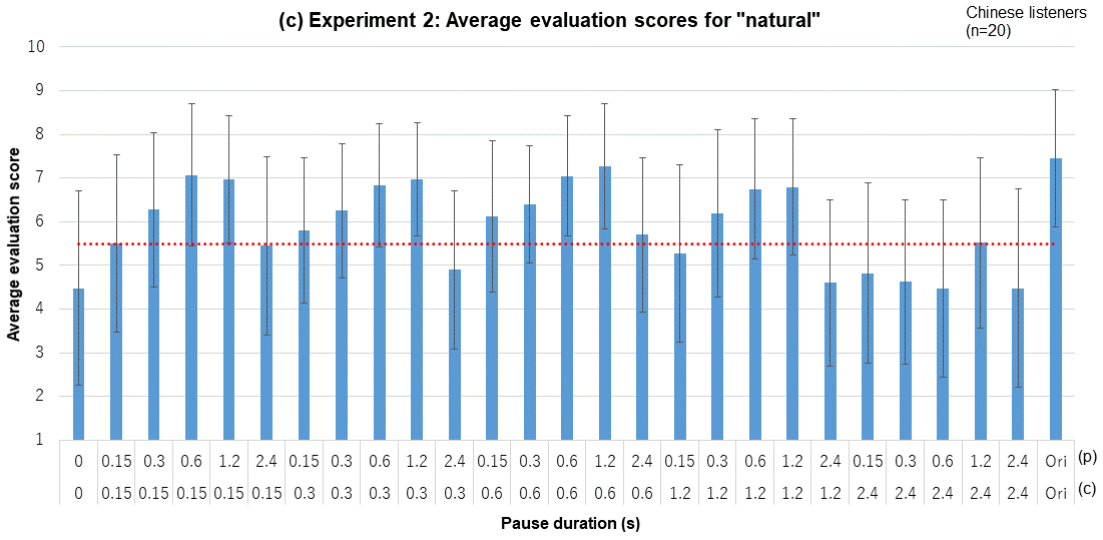

(k) Experiment 2: Average evaluation scores for "friendly"
English listeners (n=21)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
0 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 Ori (p)
0 0.15 0.15 0.15 0.15 0.15 0.3 0.3 0.3 0.3 0.3 0.6 0.6 0.6 0.6 0.6 1.2 1.2 1.2 1.2 1.2 2.4 2.4 2.4 2.4 2.4 Ori (c)
Pause duration (s)

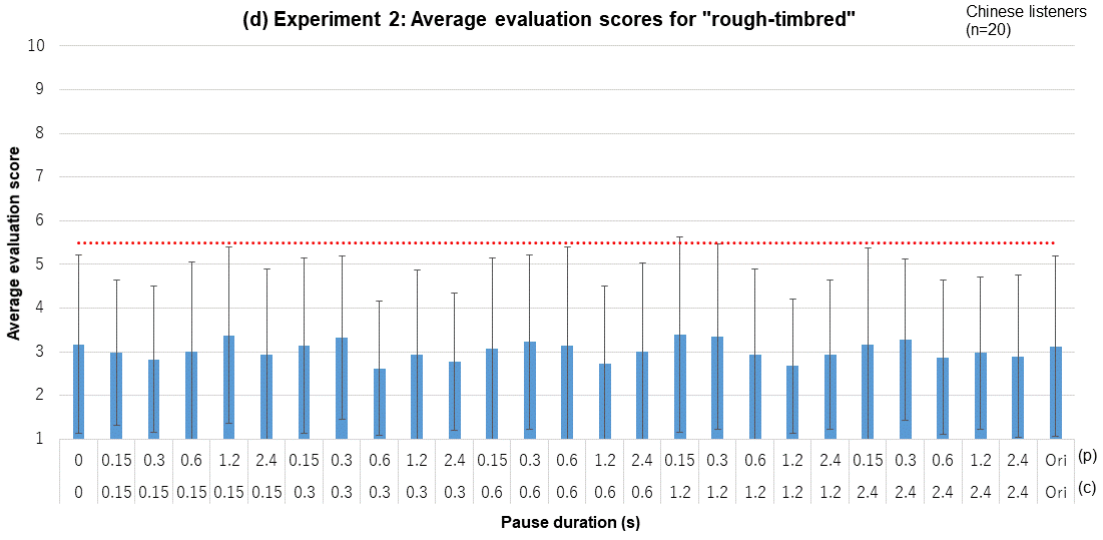


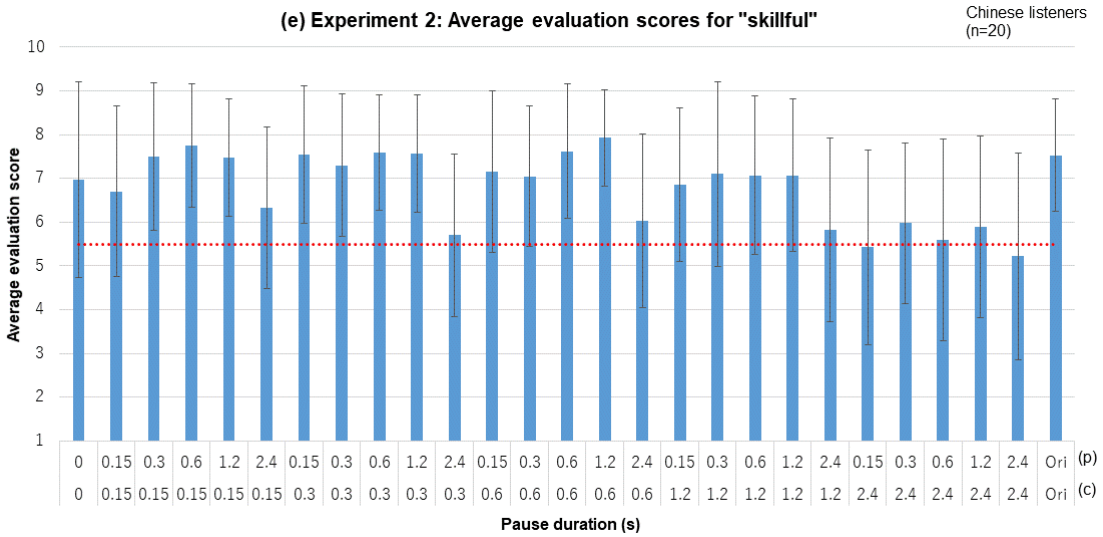
Chinese group

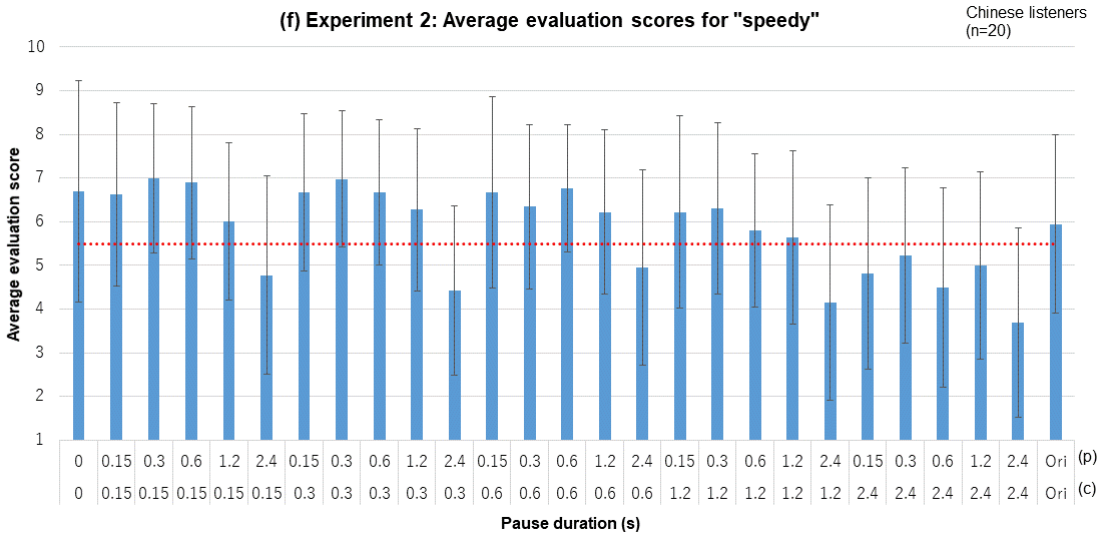


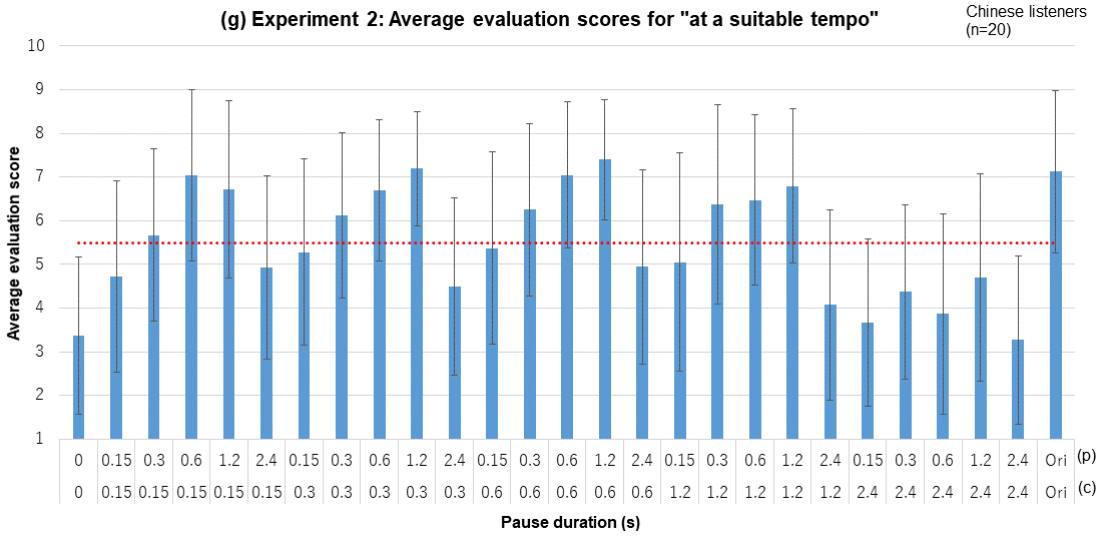

(b) Experiment 2: Average evaluation scores for "rushed"
Chinese listeners (n=20)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
0 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 Ori (p)
0 0.15 0.15 0.15 0.15 0.15 0.3 0.3 0.3 0.3 0.3 0.6 0.6 0.6 0.6 0.6 1.2 1.2 1.2 1.2 1.2 2.4 2.4 2.4 2.4 2.4 Ori (c)
Pause duration (s)

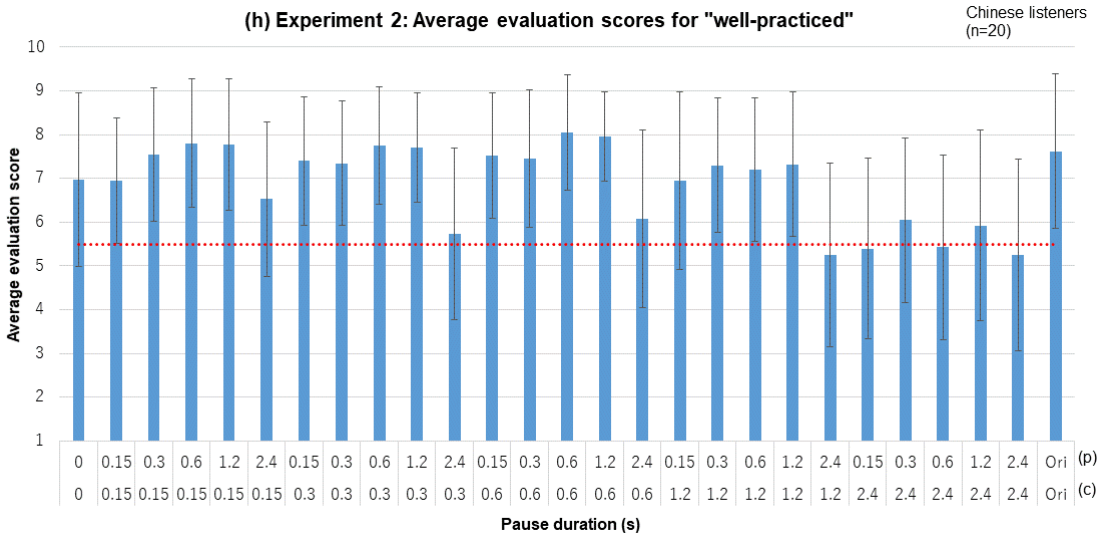

(c) Experiment 2: Average evaluation scores for "natural"
Chinese listeners (n=20)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
0 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 Ori (p)
0 0.15 0.15 0.15 0.15 0.15 0.3 0.3 0.3 0.3 0.3 0.6 0.6 0.6 0.6 0.6 1.2 1.2 1.2 1.2 1.2 2.4 2.4 2.4 2.4 2.4 Ori (c)
Pause duration (s)

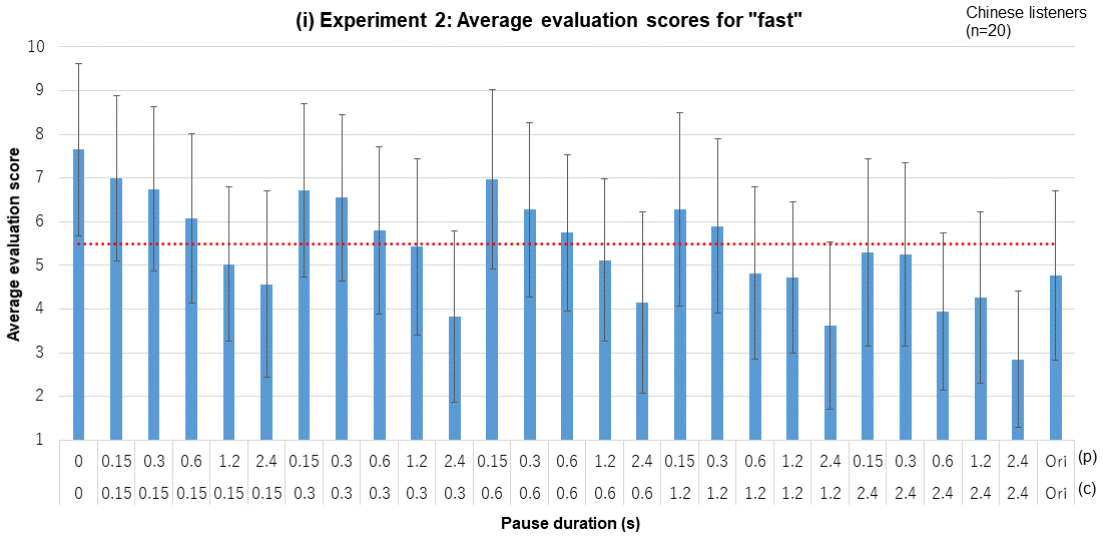

(d) Experiment 2: Average evaluation scores for "rough-timbred"
Chinese listeners (n=20)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
0 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 Ori (p)
0 0.15 0.15 0.15 0.15 0.15 0.3 0.3 0.3 0.3 0.3 0.6 0.6 0.6 0.6 0.6 1.2 1.2 1.2 1.2 1.2 2.4 2.4 2.4 2.4 2.4 Ori (c)
Pause duration (s)

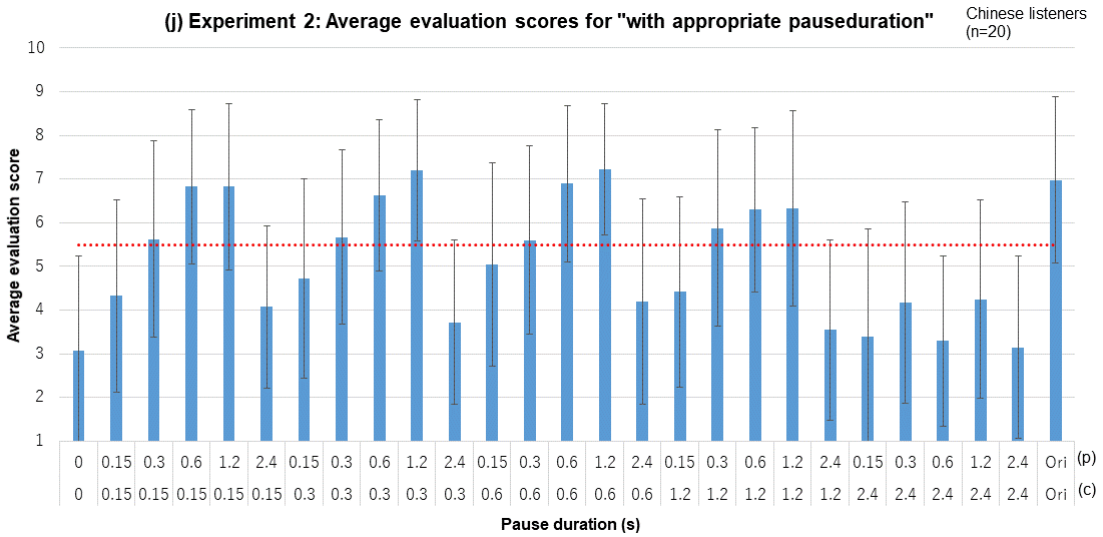

(e) Experiment 2: Average evaluation scores for "skillful"
Chinese listeners (n=20)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
0 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 Ori (p)
0 0.15 0.15 0.15 0.15 0.15 0.3 0.3 0.3 0.3 0.3 0.6 0.6 0.6 0.6 0.6 1.2 1.2 1.2 1.2 1.2 2.4 2.4 2.4 2.4 2.4 Ori (c)
Pause duration (s)

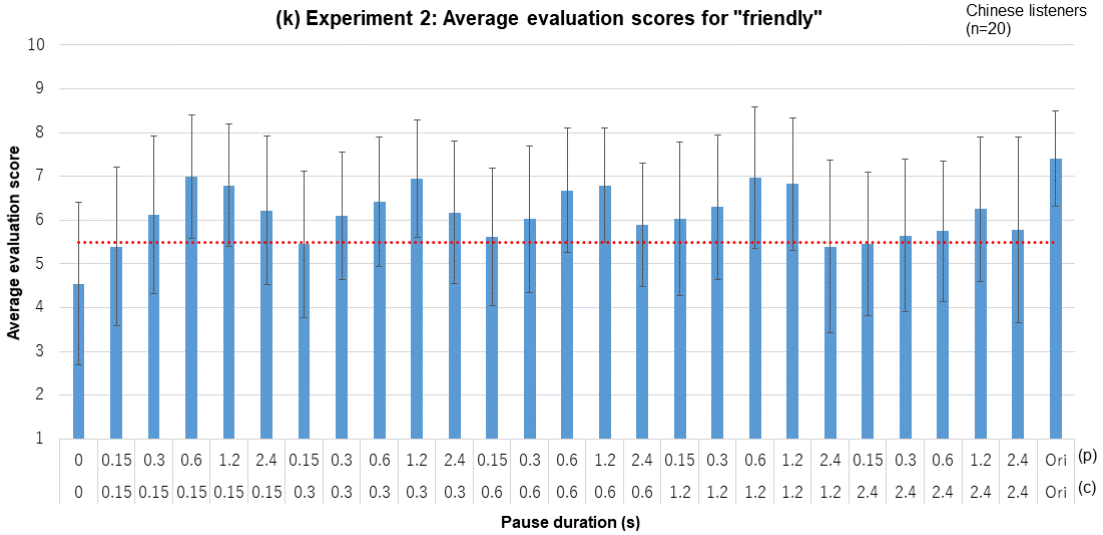

(f) Experiment 2: Average evaluation scores for "speedy"
Chinese listeners (n=20)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
0 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 Ori (p)
0 0.15 0.15 0.15 0.15 0.15 0.3 0.3 0.3 0.3 0.3 0.6 0.6 0.6 0.6 0.6 1.2 1.2 1.2 1.2 1.2 2.4 2.4 2.4 2.4 2.4 Ori (c)
Pause duration (s)

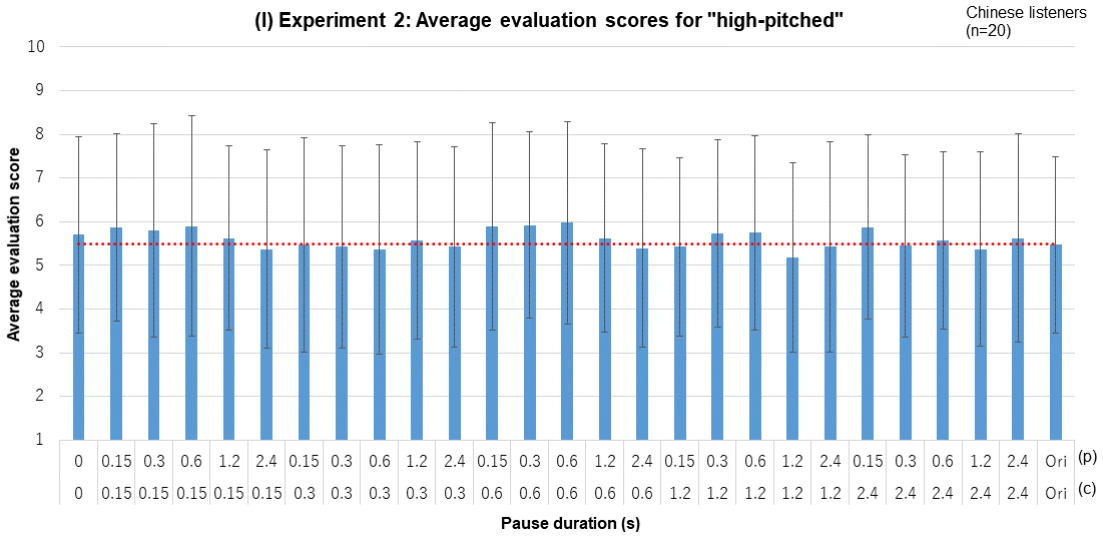

(g) Experiment 2: Average evaluation scores for "at a suitable tempo"
Chinese listeners (n=20)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
0 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 Ori (p)
0 0.15 0.15 0.15 0.15 0.15 0.3 0.3 0.3 0.3 0.3 0.6 0.6 0.6 0.6 0.6 1.2 1.2 1.2 1.2 1.2 2.4 2.4 2.4 2.4 2.4 Ori (c)
Pause duration (s)


(h) Experiment 2: Average evaluation scores for "well-practiced"
Chinese listeners (n=20)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
0 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 Ori (p)
0 0.15 0.15 0.15 0.15 0.15 0.3 0.3 0.3 0.3 0.3 0.6 0.6 0.6 0.6 0.6 1.2 1.2 1.2 1.2 1.2 2.4 2.4 2.4 2.4 2.4 Ori (c)
Pause duration (s)

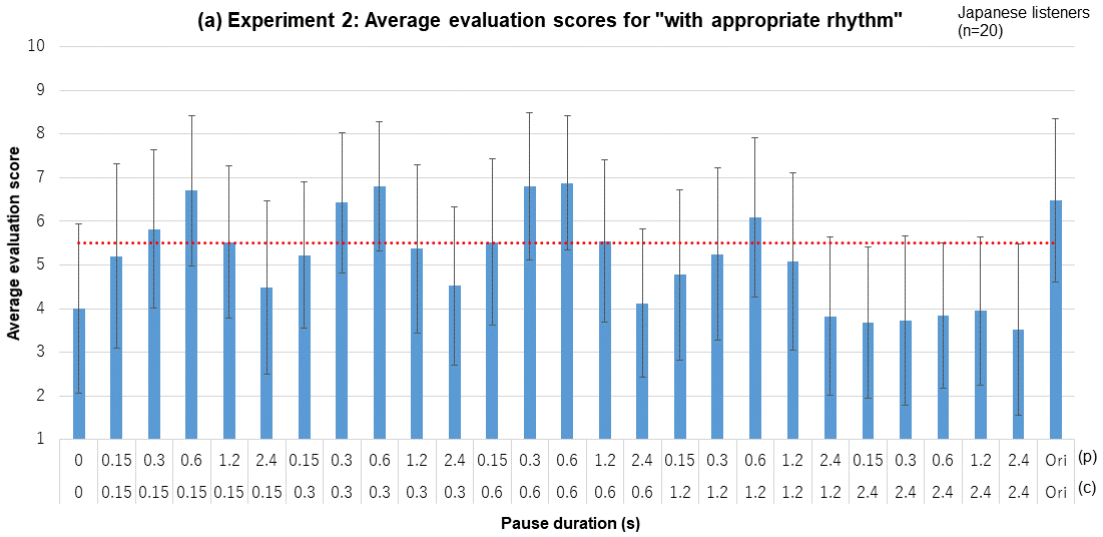

(i) Experiment 2: Average evaluation scores for "fast"
Chinese listeners (n=20)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
0 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 Ori (p)
0 0.15 0.15 0.15 0.15 0.15 0.3 0.3 0.3 0.3 0.3 0.6 0.6 0.6 0.6 0.6 1.2 1.2 1.2 1.2 1.2 2.4 2.4 2.4 2.4 2.4 Ori (c)
Pause duration (s)

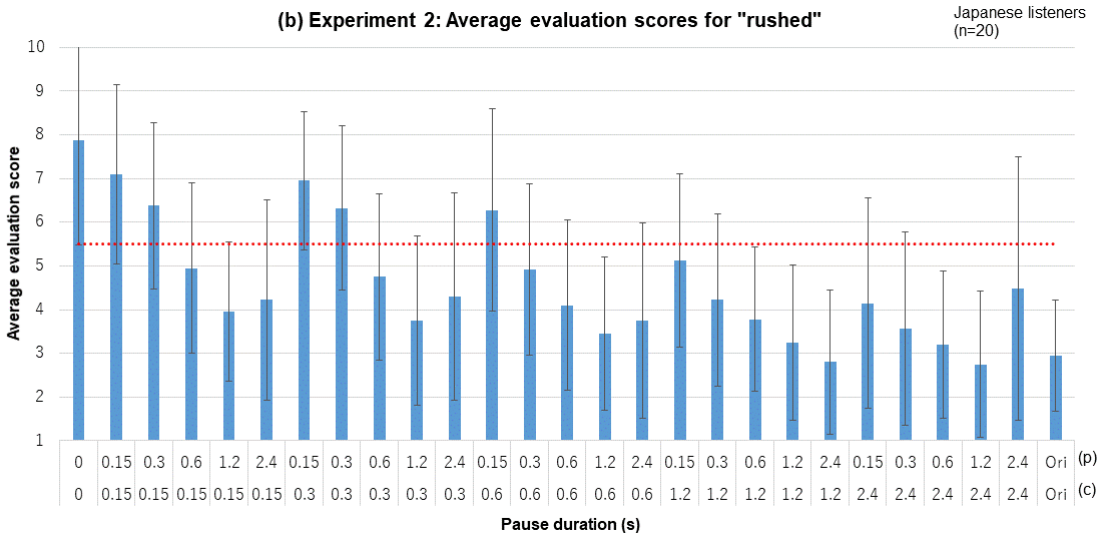

(j) Experiment 2: Average evaluation scores for "with appropriate pauseduration"
Chinese listeners (n=20)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
0 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 Ori (p)
0 0.15 0.15 0.15 0.15 0.15 0.3 0.3 0.3 0.3 0.3 0.6 0.6 0.6 0.6 0.6 1.2 1.2 1.2 1.2 1.2 2.4 2.4 2.4 2.4 2.4 Ori (c)
Pause duration (s)

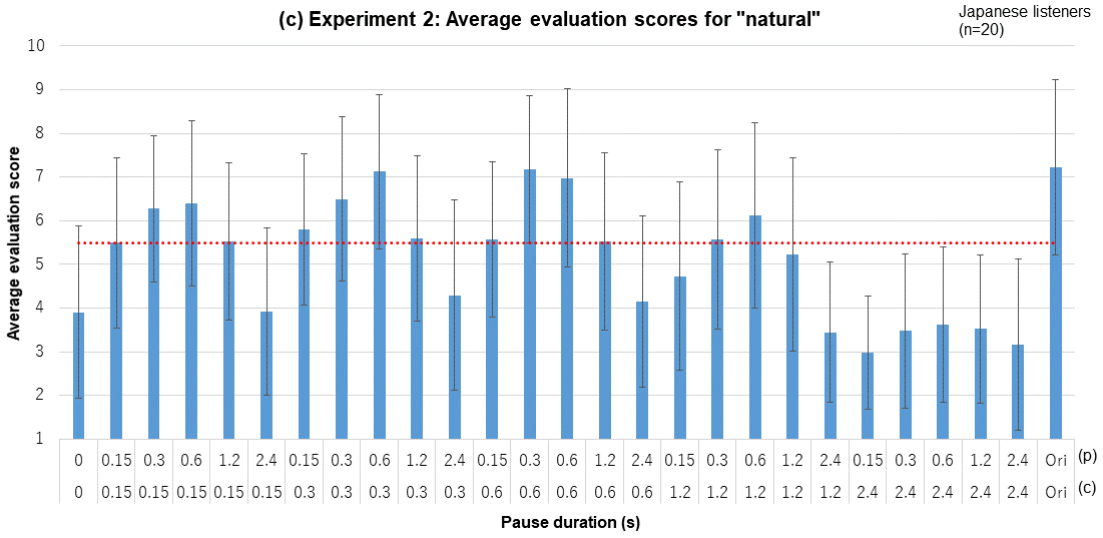

(k) Experiment 2: Average evaluation scores for "friendly"
Chinese listeners (n=20)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
0 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 Ori (p)
0 0.15 0.15 0.15 0.15 0.15 0.3 0.3 0.3 0.3 0.3 0.6 0.6 0.6 0.6 0.6 1.2 1.2 1.2 1.2 1.2 2.4 2.4 2.4 2.4 2.4 Ori (c)
Pause duration (s)

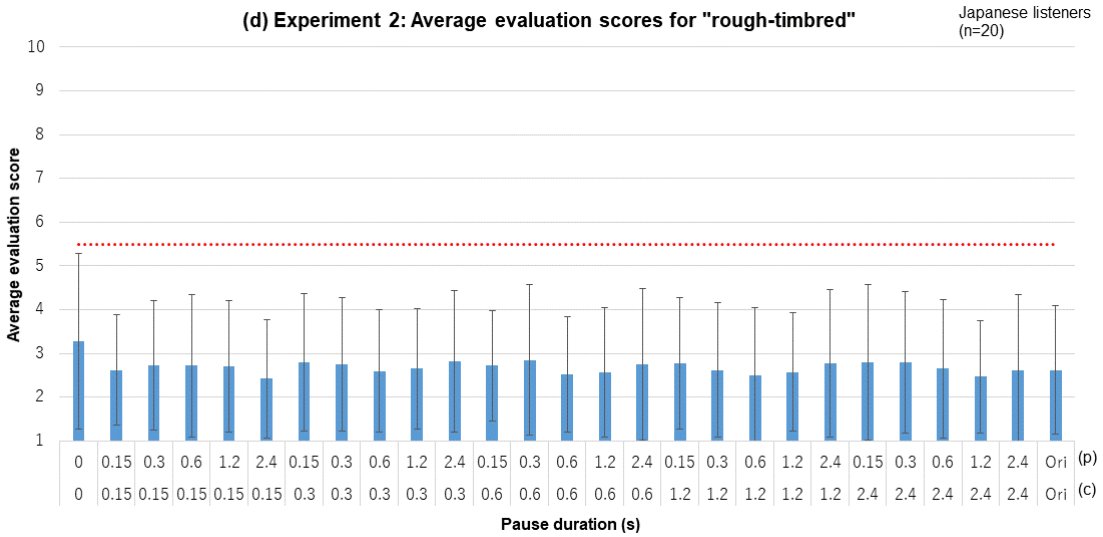

(l) Experiment 2: Average evaluation scores for "high-pitched"
Chinese listeners (n=20)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
0 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 Ori (p)
0 0.15 0.15 0.15 0.15 0.15 0.3 0.3 0.3 0.3 0.3 0.6 0.6 0.6 0.6 0.6 1.2 1.2 1.2 1.2 1.2 2.4 2.4 2.4 2.4 2.4 Ori (c)
Pause duration (s)

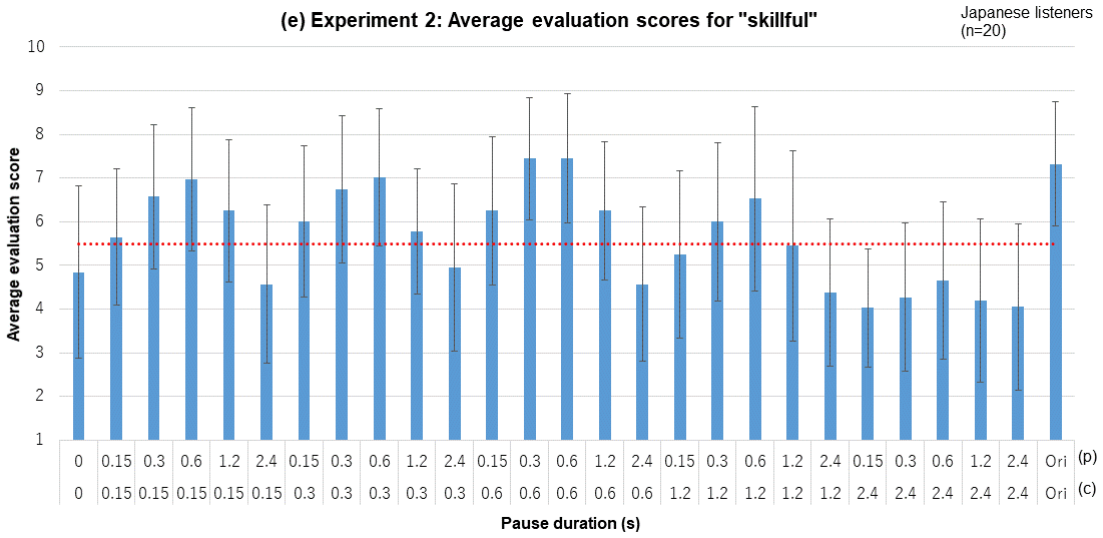
Japanese group

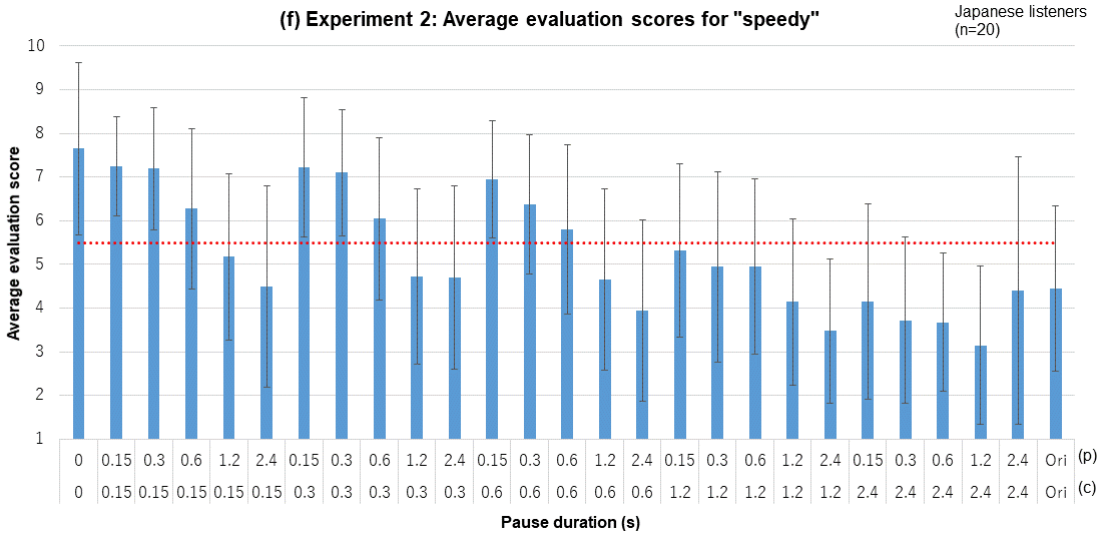


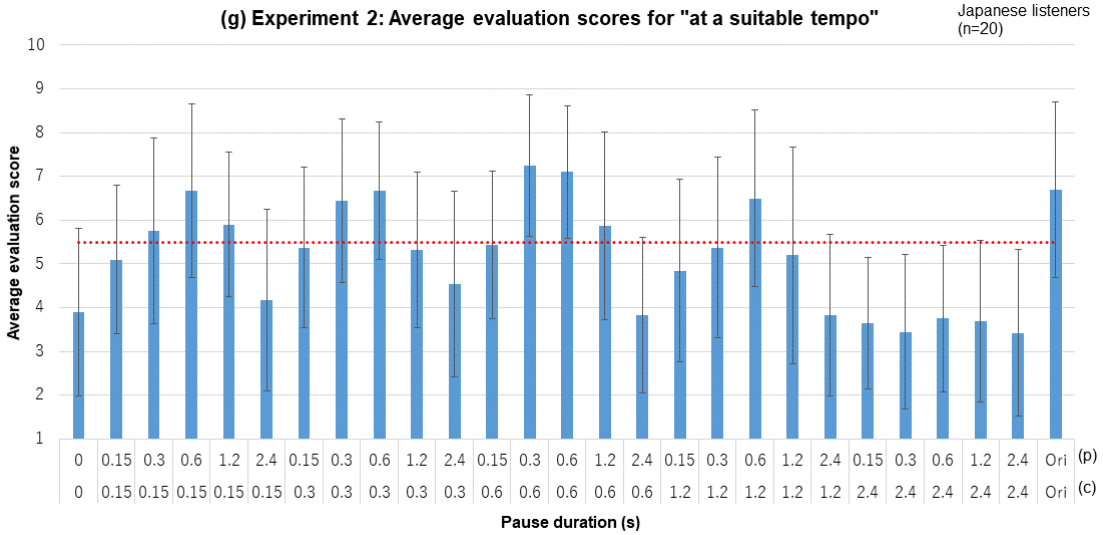


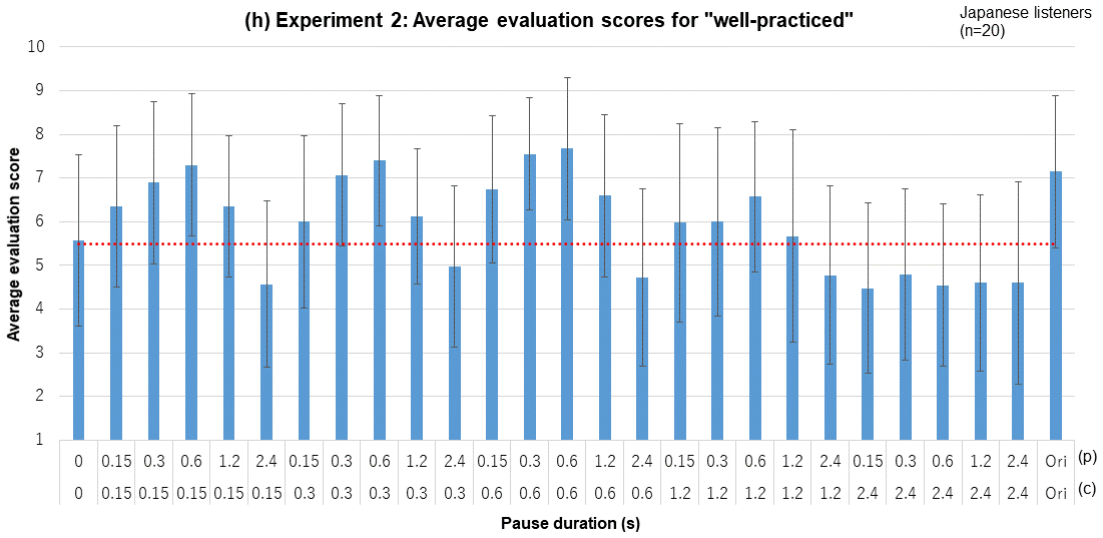


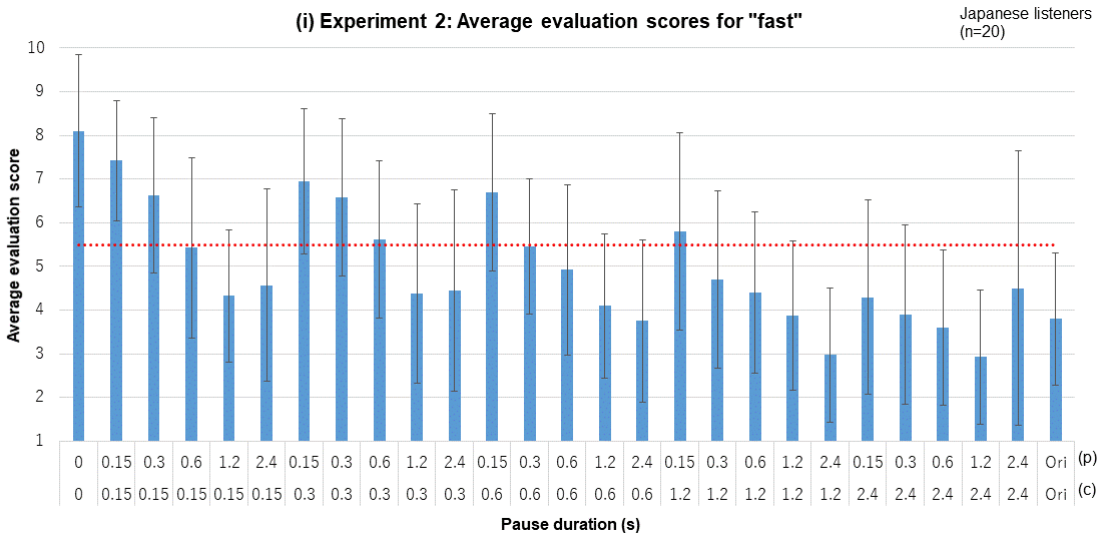

(d) Experiment 2: Average evaluation scores for "rough-timbred"
Japanese listeners (n=20)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
0 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 Ori (p)
0 0.15 0.15 0.15 0.15 0.15 0.3 0.3 0.3 0.3 0.3 0.6 0.6 0.6 0.6 0.6 1.2 1.2 1.2 1.2 1.2 2.4 2.4 2.4 2.4 2.4 Ori (c)
Pause duration (s)

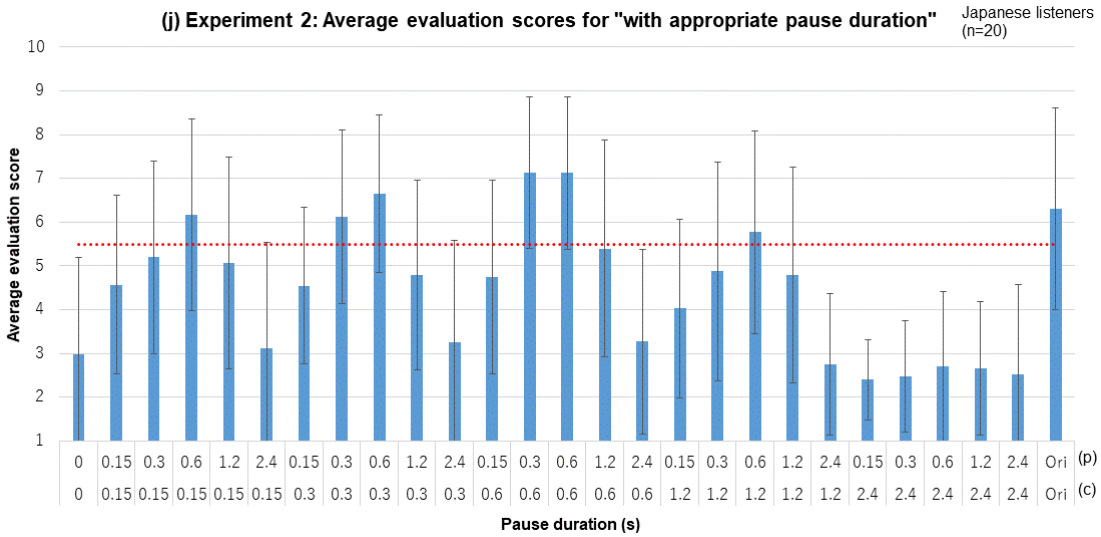

(e) Experiment 2: Average evaluation scores for "skillful"
Japanese listeners (n=20)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
0 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 Ori (p)
0 0.15 0.15 0.15 0.15 0.15 0.3 0.3 0.3 0.3 0.3 0.6 0.6 0.6 0.6 0.6 1.2 1.2 1.2 1.2 1.2 2.4 2.4 2.4 2.4 2.4 Ori (c)
Pause duration (s)

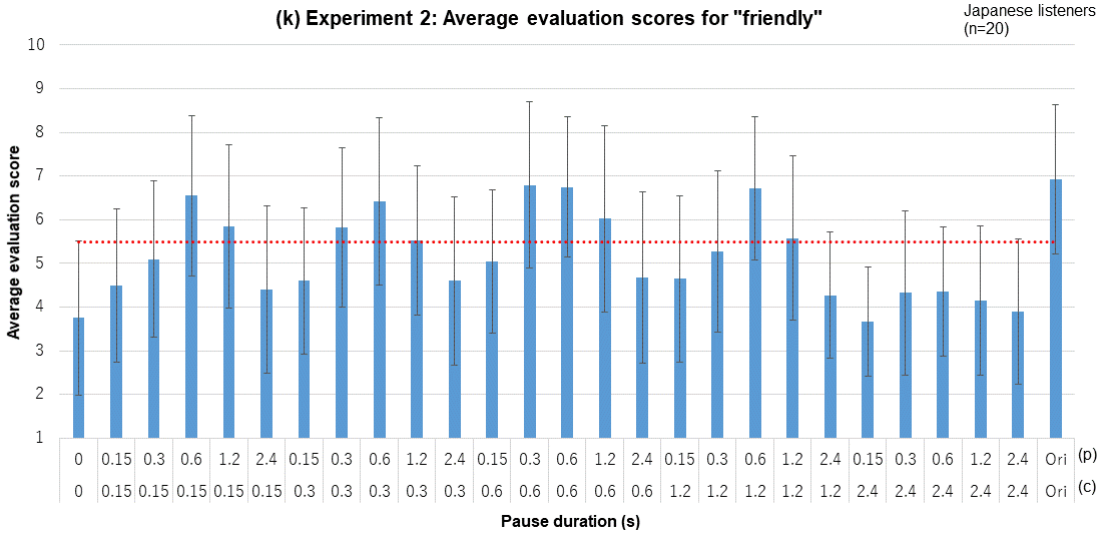

(f) Experiment 2: Average evaluation scores for "speedy"
Japanese listeners (n=20)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
0 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 Ori (p)
0 0.15 0.15 0.15 0.15 0.15 0.3 0.3 0.3 0.3 0.3 0.6 0.6 0.6 0.6 0.6 1.2 1.2 1.2 1.2 1.2 2.4 2.4 2.4 2.4 2.4 Ori (c)
Pause duration (s)

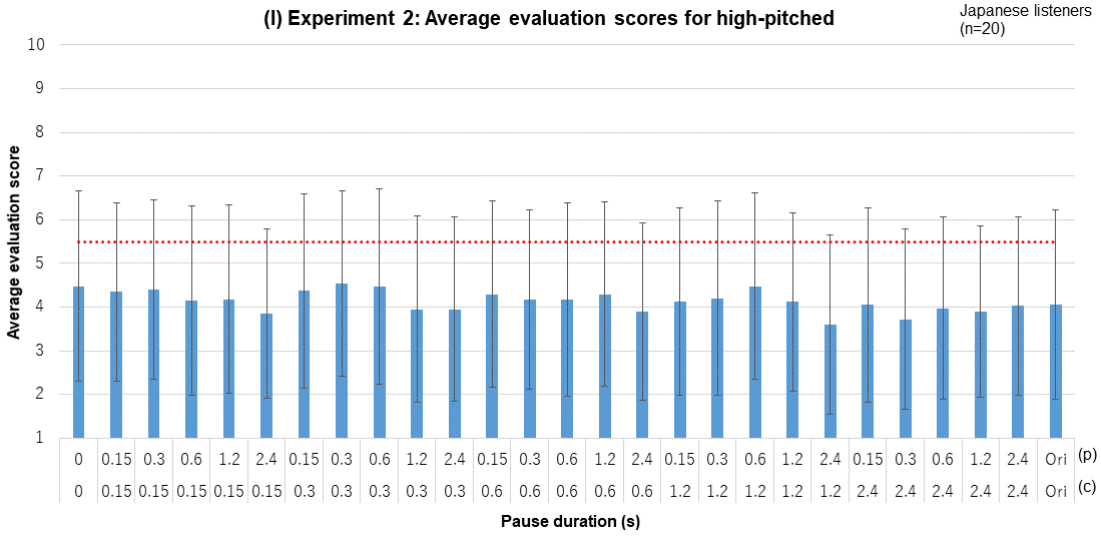

(g) Experiment 2: Average evaluation scores for "at a suitable tempo"
Japanese listeners (n=20)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
0 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 Ori (p)
0 0.15 0.15 0.15 0.15 0.15 0.3 0.3 0.3 0.3 0.3 0.6 0.6 0.6 0.6 0.6 1.2 1.2 1.2 1.2 1.2 2.4 2.4 2.4 2.4 2.4 Ori (c)
Pause duration (s)


(h) Experiment 2: Average evaluation scores for "well-practiced"
Japanese listeners (n=20)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
0 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 Ori (p)
0 0.15 0.15 0.15 0.15 0.15 0.3 0.3 0.3 0.3 0.3 0.6 0.6 0.6 0.6 0.6 1.2 1.2 1.2 1.2 1.2 2.4 2.4 2.4 2.4 2.4 Ori (c)
Pause duration (s)


(i) Experiment 2: Average evaluation scores for "fast"
Japanese listeners (n=20)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
0 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 Ori (p)
0 0.15 0.15 0.15 0.15 0.15 0.3 0.3 0.3 0.3 0.3 0.6 0.6 0.6 0.6 0.6 1.2 1.2 1.2 1.2 1.2 2.4 2.4 2.4 2.4 2.4 Ori (c)
Pause duration (s)

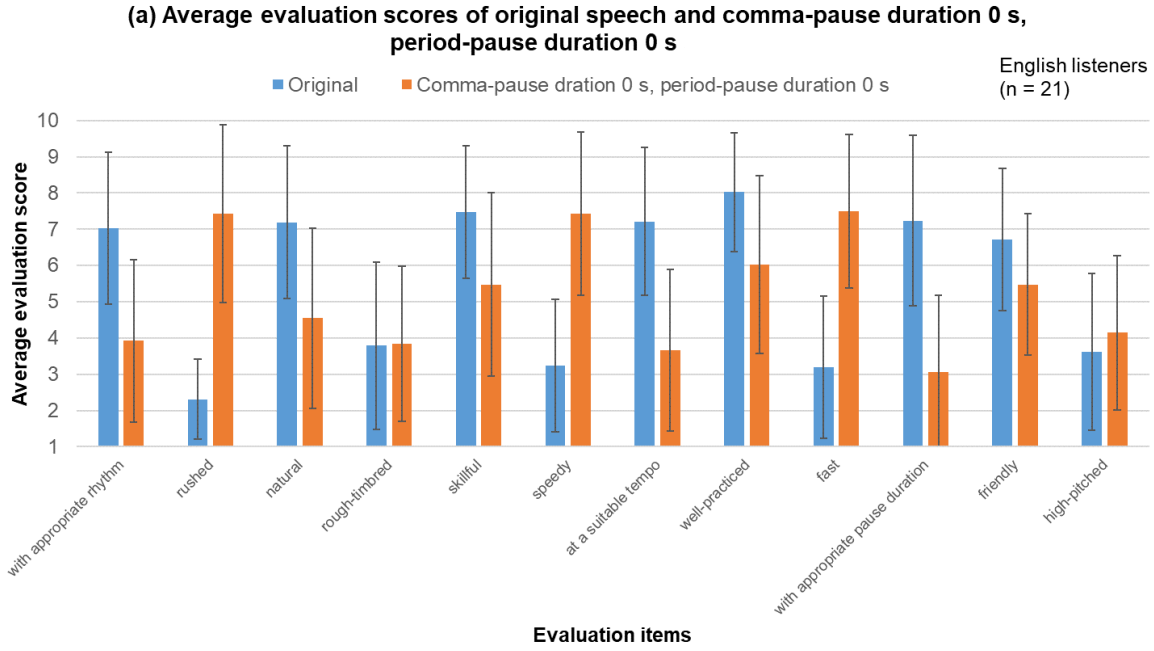

(j) Experiment 2: Average evaluation scores for "with appropriate pause duration"
Japanese listeners (n=20)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
0 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 Ori (p)
0 0.15 0.15 0.15 0.15 0.15 0.3 0.3 0.3 0.3 0.3 0.6 0.6 0.6 0.6 0.6 1.2 1.2 1.2 1.2 1.2 2.4 2.4 2.4 2.4 2.4 Ori (c)
Pause duration (s)

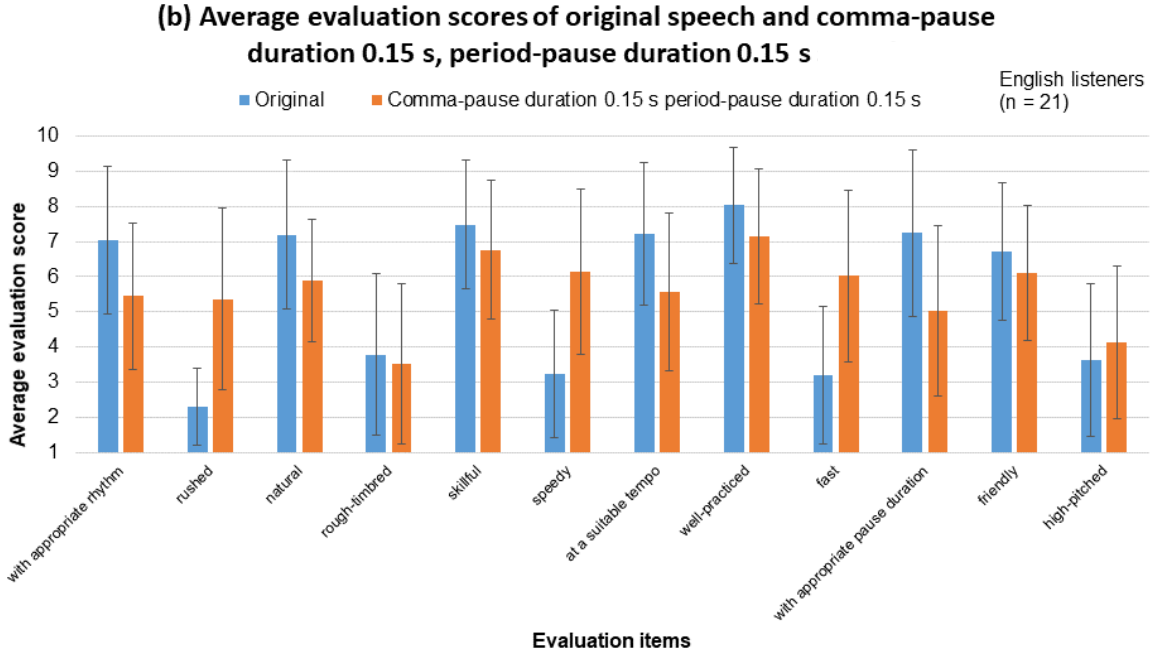

(k) Experiment 2: Average evaluation scores for "friendly"
Japanese listeners (n=20)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
0 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 Ori (p)
0 0.15 0.15 0.15 0.15 0.15 0.3 0.3 0.3 0.3 0.3 0.6 0.6 0.6 0.6 0.6 1.2 1.2 1.2 1.2 1.2 2.4 2.4 2.4 2.4 2.4 Ori (c)
Pause duration (s)

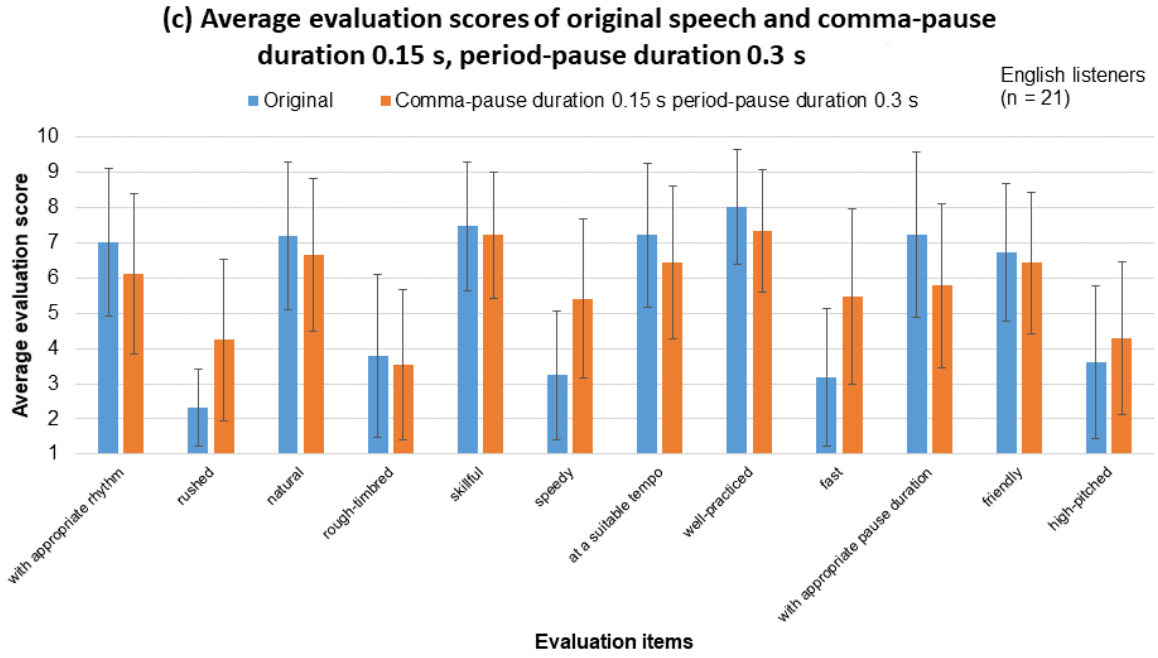

(l) Experiment 2: Average evaluation scores for high-pitched
Japanese listeners (n=20)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
0 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 0.15 0.3 0.6 1.2 2.4 Ori (p)
0 0.15 0.15 0.15 0.15 0.15 0.3 0.3 0.3 0.3 0.3 0.6 0.6 0.6 0.6 0.6 1.2 1.2 1.2 1.2 1.2 2.4 2.4 2.4 2.4 2.4 Ori (c)
Pause duration (s)

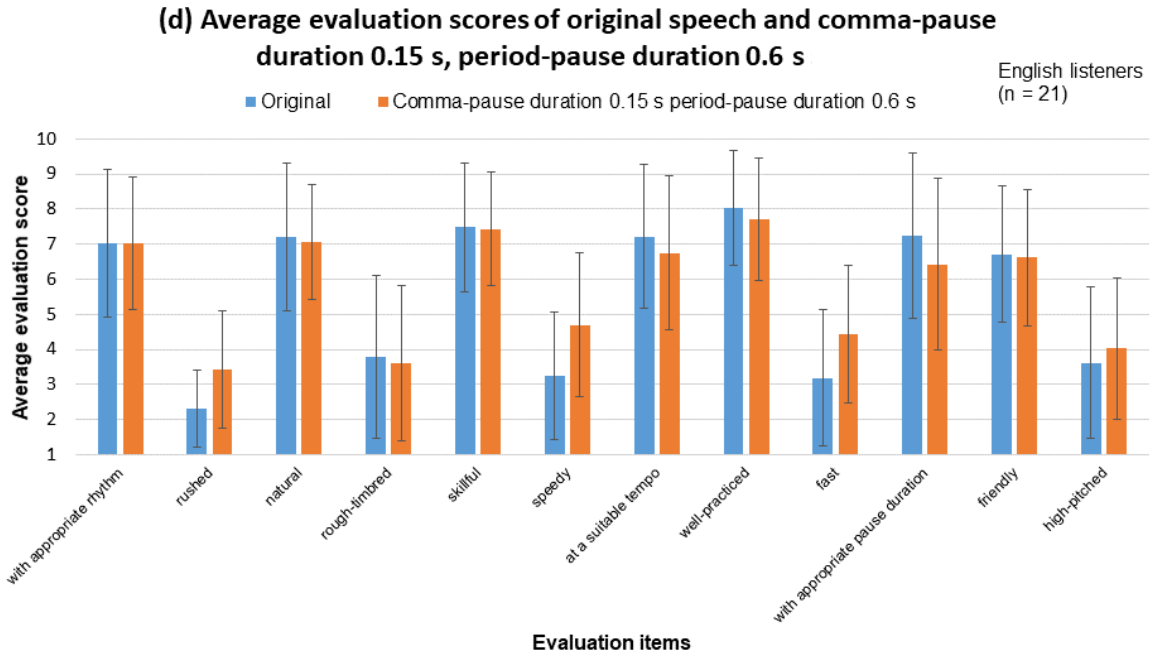
Appendix K. Average evaluation scores compared to original speech

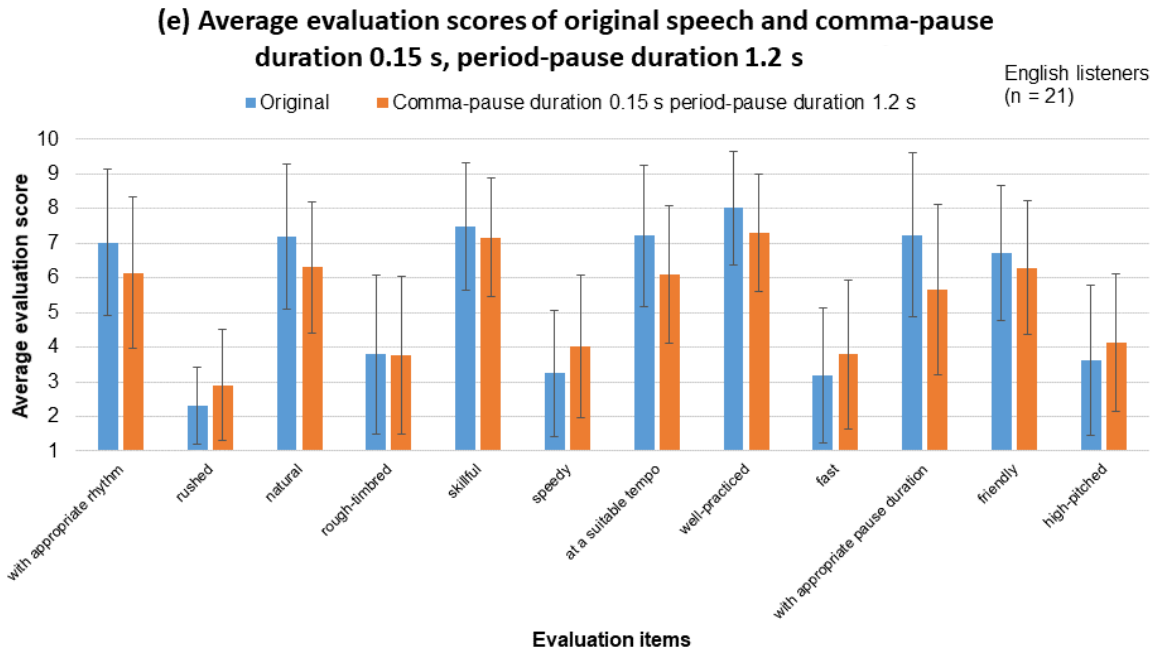
English group

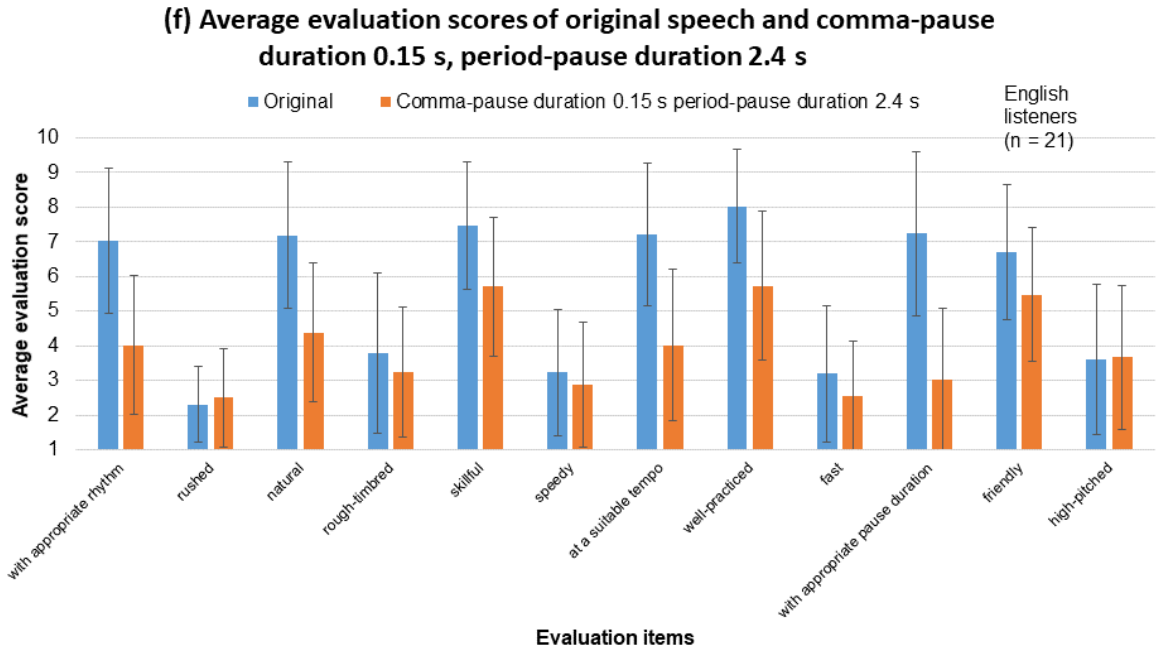


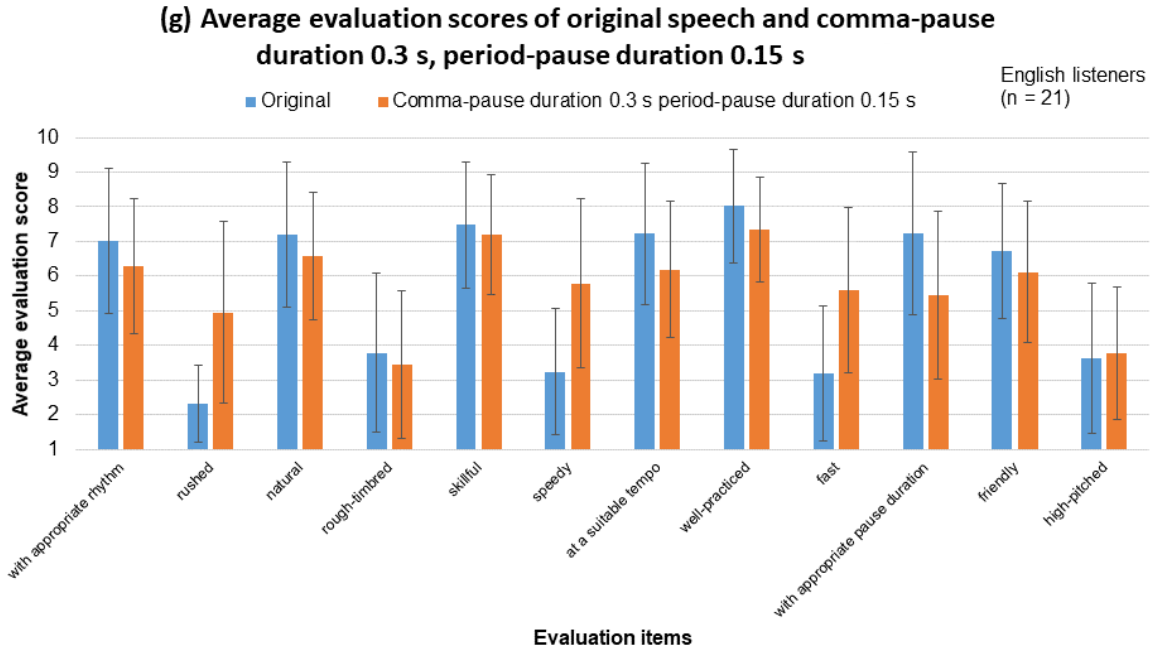


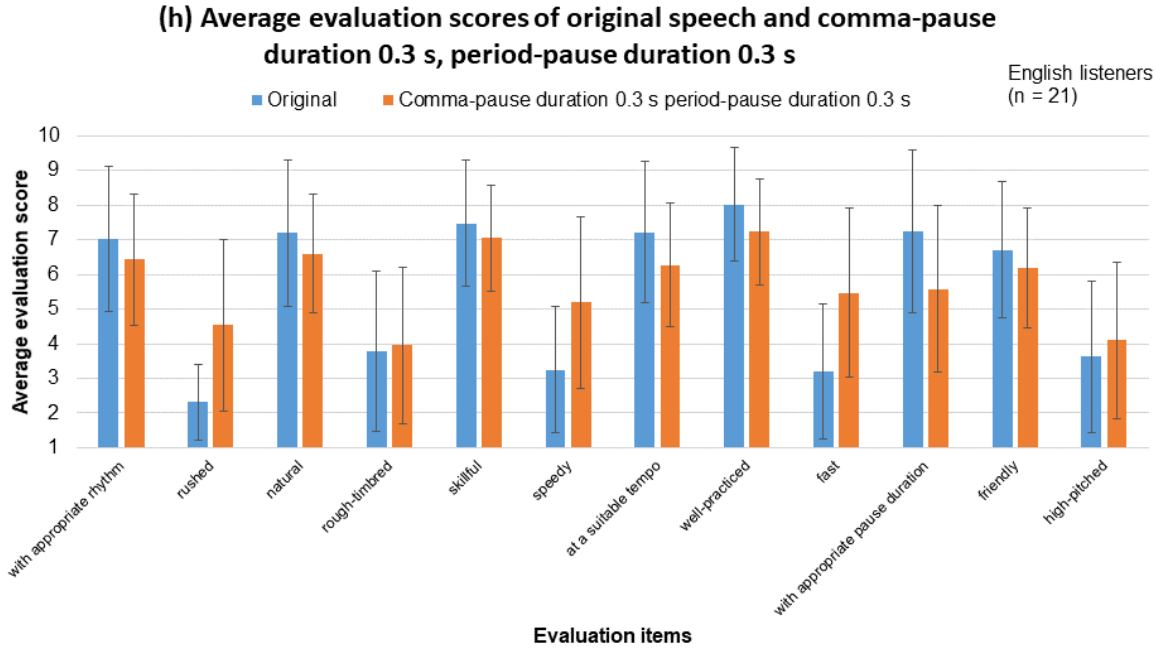

(c) Average evaluation scores of original speech and comma-pause duration 0.15 s, period-pause duration 0.3 s
English listeners (n = 21)
Original
Comma-pause duration 0.15 s period-pause duration 0.3 s
Average evaluation score
10
9
8
7
6
5
4
3
2
1
with appropriate rhythm
rushed
natural
rough-timbred
skillful
speedy
at a suitable tempo
well-practiced
fast
with appropriate pause duration
friendly
high-pitched
Evaluation items

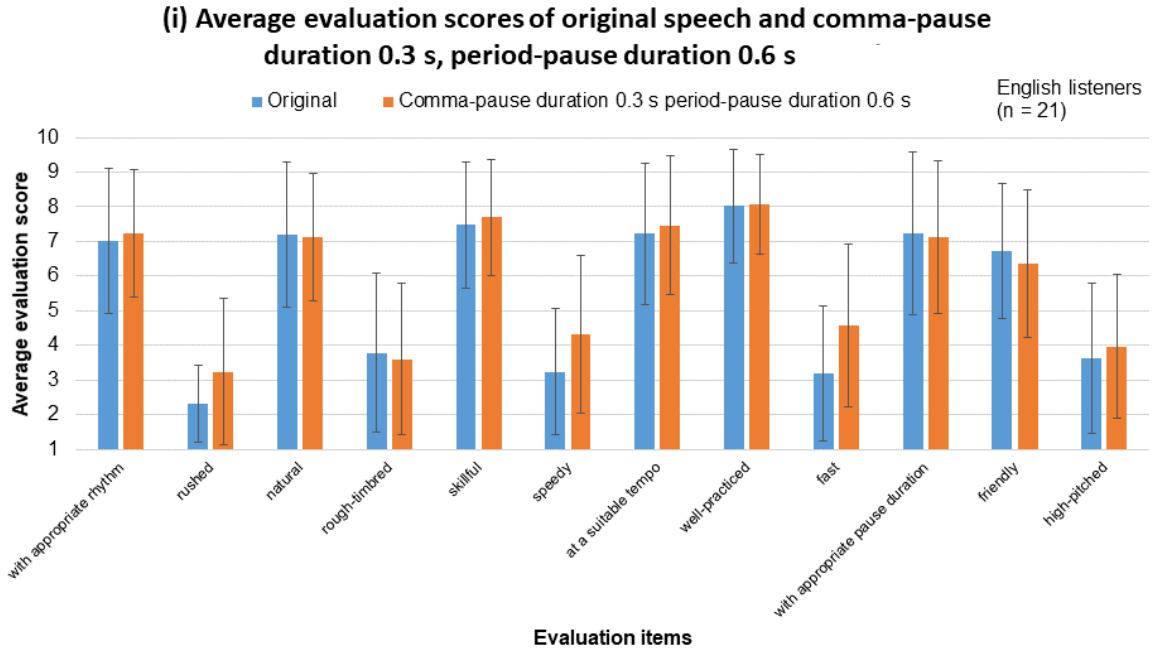

(d) Average evaluation scores of original speech and comma-pause duration 0.15 s, period-pause duration 0.6 s
English listeners (n = 21)
Original
Comma-pause duration 0.15 s period-pause duration 0.6 s
Average evaluation score
10
9
8
7
6
5
4
3
2
1
with appropriate rhythm
rushed
natural
rough-timbred
skillful
speedy
at a suitable tempo
well-practiced
fast
with appropriate pause duration
friendly
high-pitched
Evaluation items

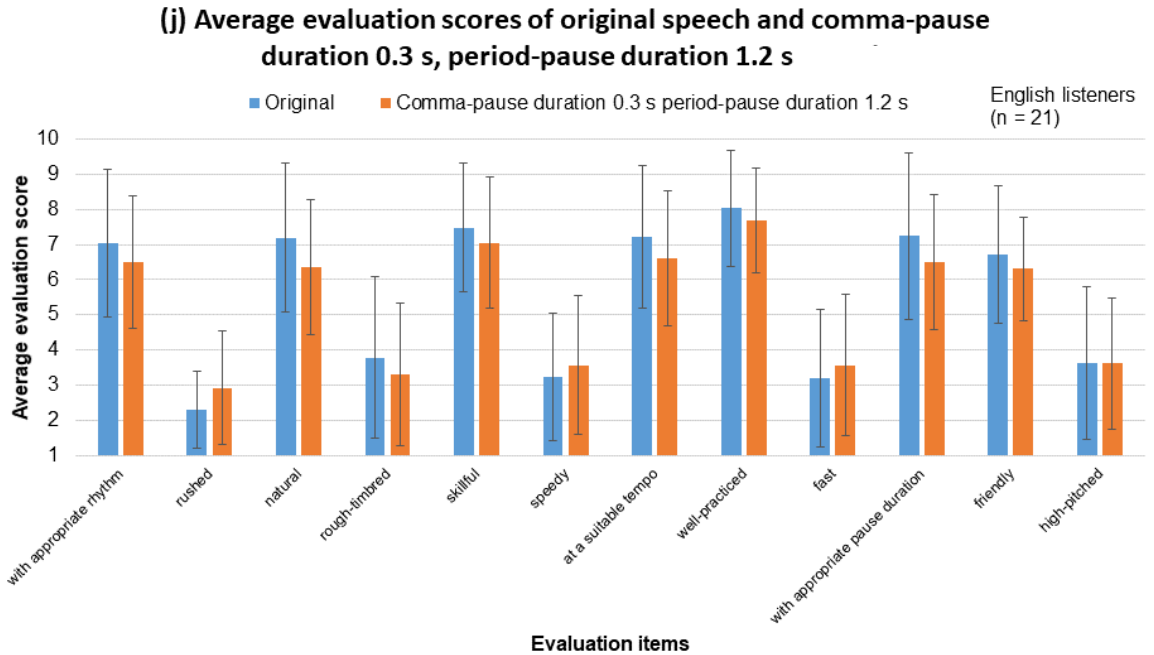

(e) Average evaluation scores of original speech and comma-pause duration 0.15 s, period-pause duration 1.2 s
English listeners (n = 21)
Original
Comma-pause duration 0.15 s period-pause duration 1.2 s
Average evaluation score
10
9
8
7
6
5
4
3
2
1
with appropriate rhythm
rushed
natural
rough-timbred
skillful
speedy
at a suitable tempo
well-practiced
fast
with appropriate pause duration
friendly
high-pitched
Evaluation items


(f) Average evaluation scores of original speech and comma-pause duration 0.15 s, period-pause duration 2.4 s
English listeners (n = 21)
Original
Comma-pause duration 0.15 s period-pause duration 2.4 s
Average evaluation score
10
9
8
7
6
5
4
3
2
1
with appropriate rhythm
rushed
natural
rough-timbred
skillful
speedy
at a suitable tempo
well-practiced
fast
with appropriate pause duration
friendly
high-pitched
Evaluation items

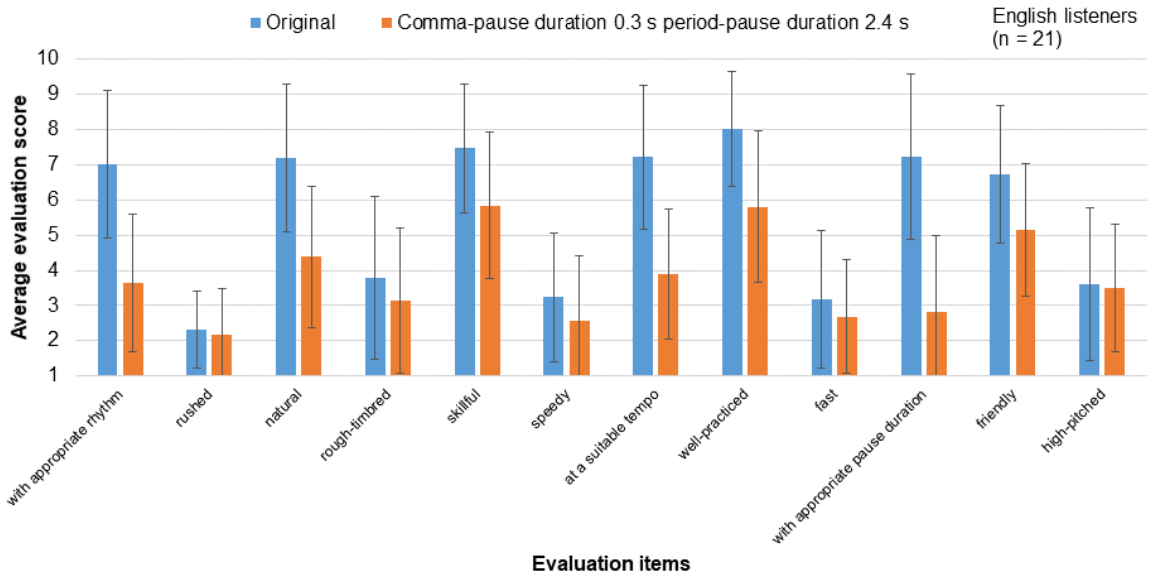

(g) Average evaluation scores of original speech and comma-pause duration 0.3 s, period-pause duration 0.15 s
English listeners (n = 21)
Original
Comma-pause duration 0.3 s period-pause duration 0.15 s
Average evaluation score
10
9
8
7
6
5
4
3
2
1
with appropriate rhythm
rushed
natural
rough-timbred
skillful
speedy
at a suitable tempo
well-practiced
fast
with appropriate pause duration
friendly
high-pitched
Evaluation items


(h) Average evaluation scores of original speech and comma-pause duration 0.3 s, period-pause duration 0.3 s
English listeners (n = 21)
Original
Comma-pause duration 0.3 s period-pause duration 0.3 s
Average evaluation score
10
9
8
7
6
5
4
3
2
1
with appropriate rhythm
rushed
natural
rough-timbred
skillful
speedy
at a suitable tempo
well-practiced
fast
with appropriate pause duration
friendly
high-pitched
Evaluation items

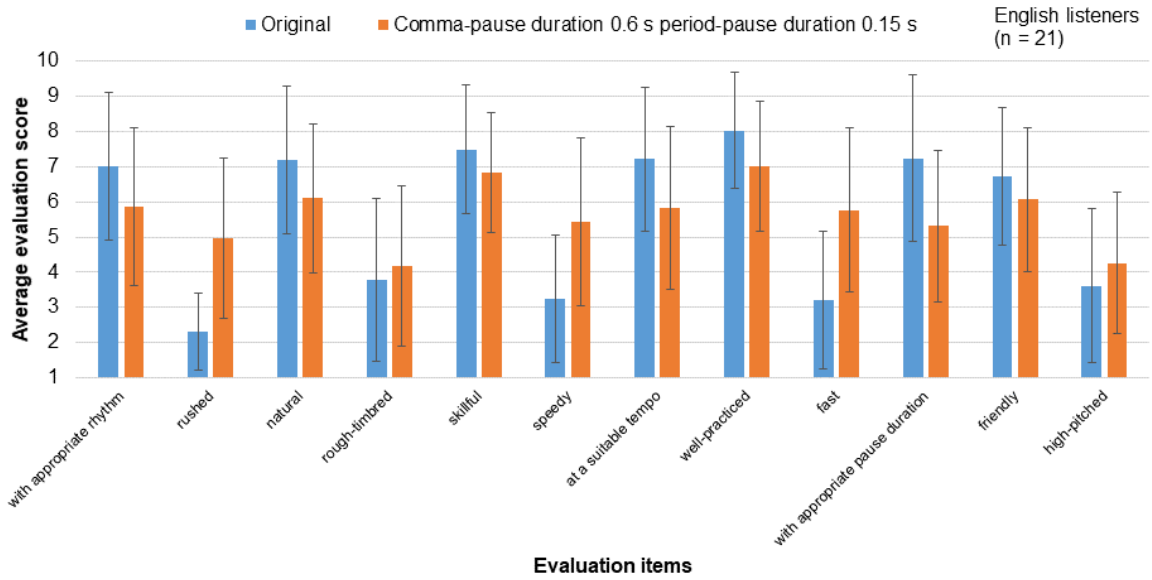

(i) Average evaluation scores of original speech and comma-pause duration 0.3 s, period-pause duration 0.6 s
Original
Comma-pause duration 0.3 s period-pause duration 0.6 s
English listeners (n = 21)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
with appropriate rhythm
rushed
natural
rough-timbred
skillful
speedy
at a suitable tempo
well-practiced
fast
with appropriate pause duration
friendly
high-pitched
Evaluation items

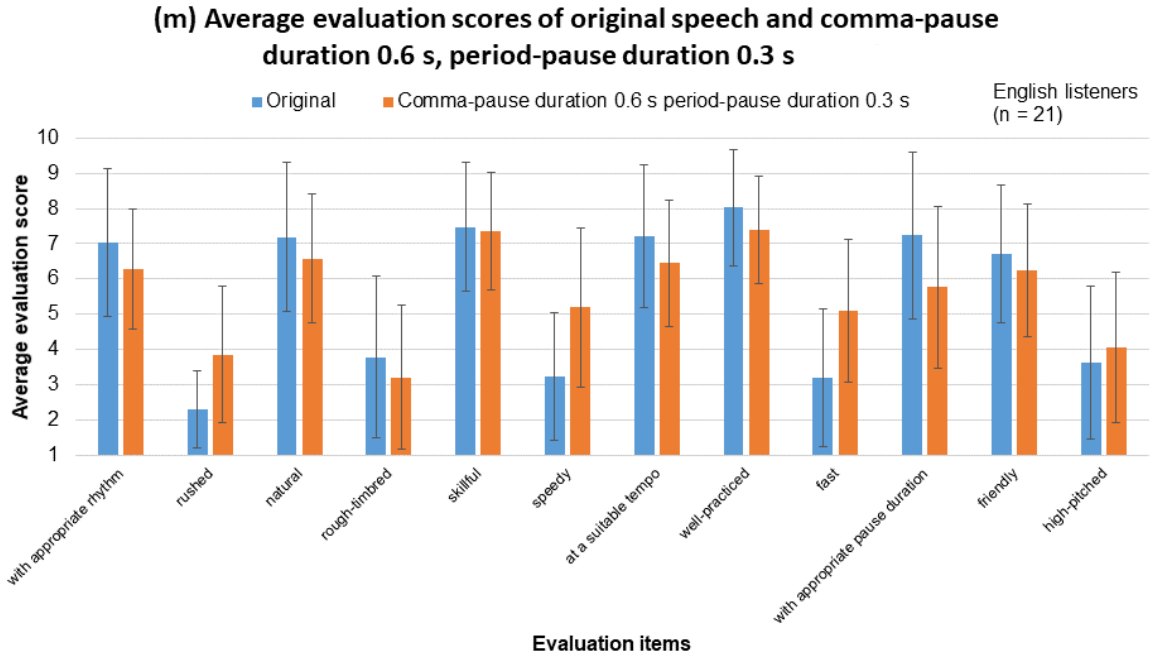

(j) Average evaluation scores of original speech and comma-pause duration 0.3 s, period-pause duration 1.2 s
Original
Comma-pause duration 0.3 s period-pause duration 1.2 s
English listeners (n = 21)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
with appropriate rhythm
rushed
natural
rough-timbred
skillful
speedy
at a suitable tempo
well-practiced
fast
with appropriate pause duration
friendly
high-pitched
Evaluation items

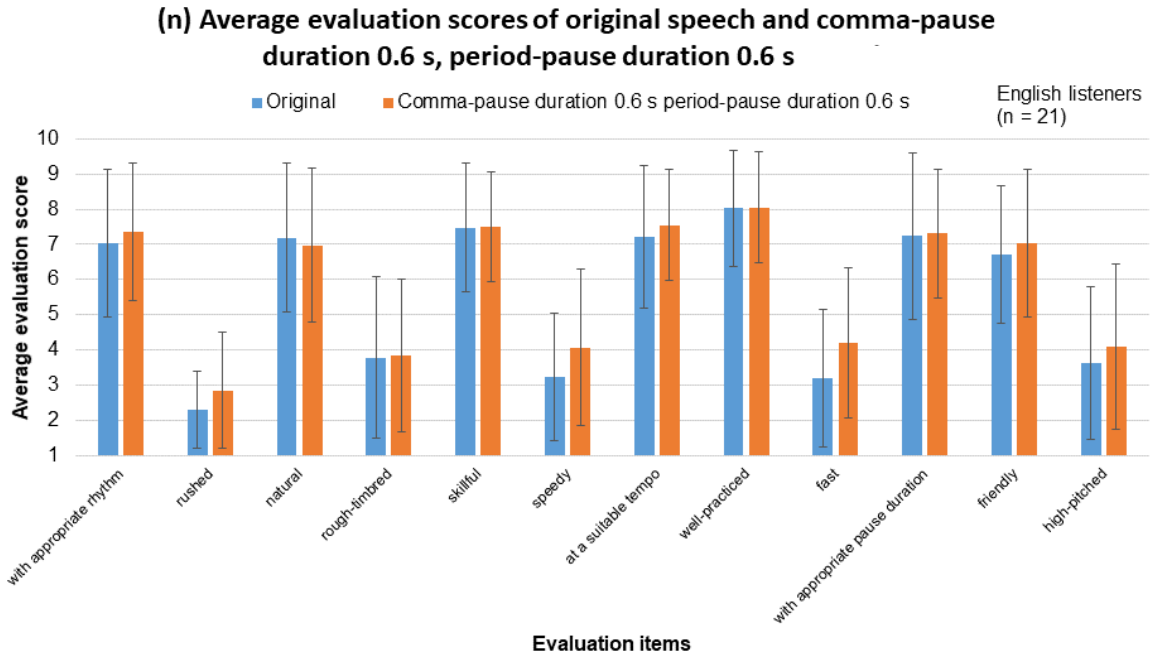
(k) Average evaluation scores of original speech and comma-pause duration 0.3 s, period-pause duration 2.4 s



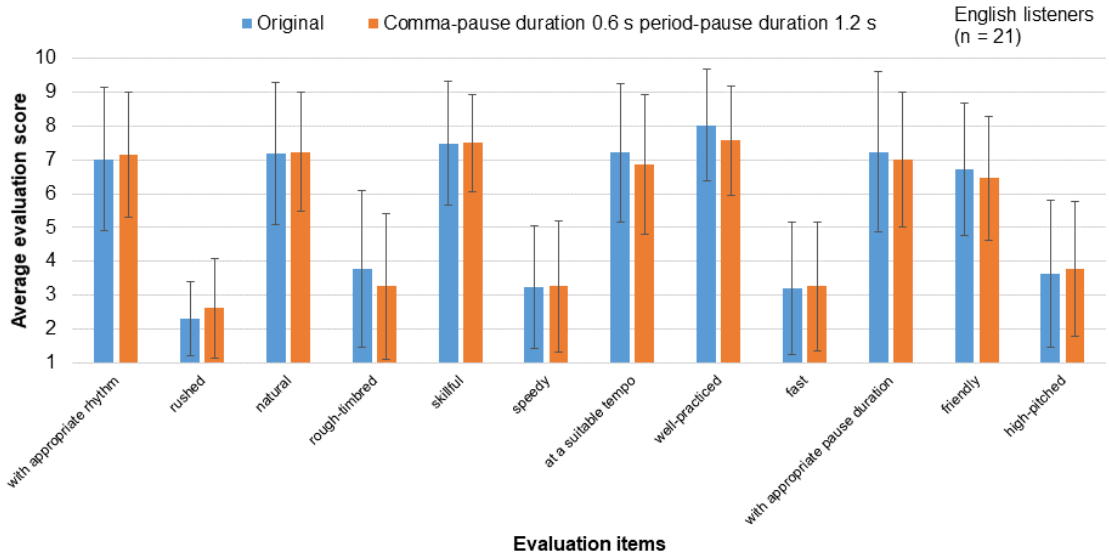
(l) Average evaluation scores of original speech and comma-pause duration 0.6 s, period-pause duration 0.15 s



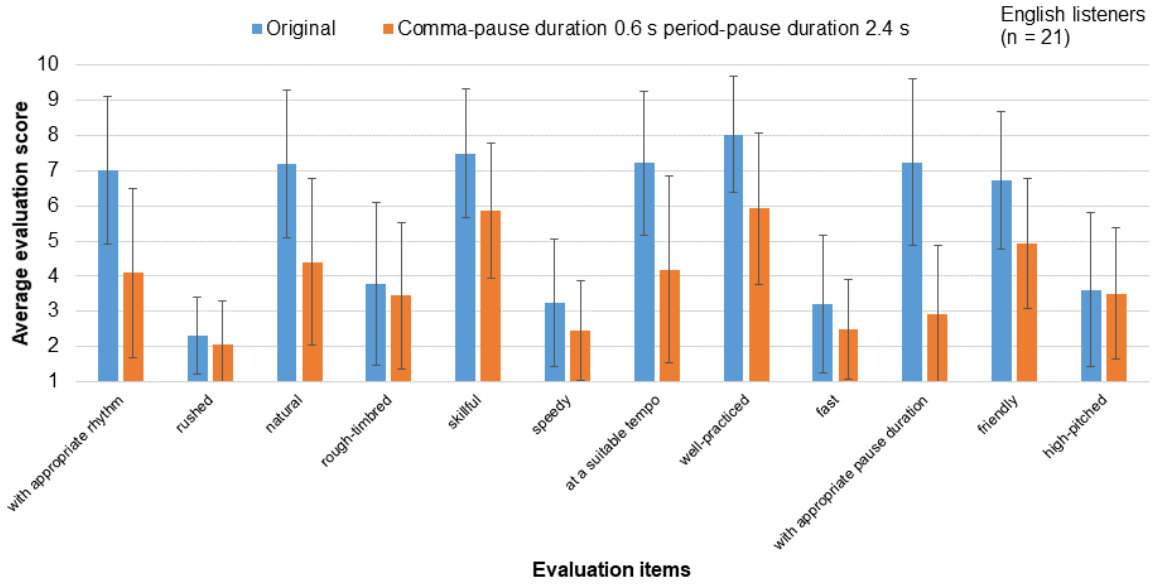

(m) Average evaluation scores of original speech and comma-pause duration 0.6 s, period-pause duration 0.3 s
Original
Comma-pause duration 0.6 s period-pause duration 0.3 s
English listeners (n = 21)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
with appropriate rhythm
rushed
natural
rough-timbred
skillful
speedy
at a suitable tempo
well-practiced
fast
with appropriate pause duration
friendly
high-pitched
Evaluation items

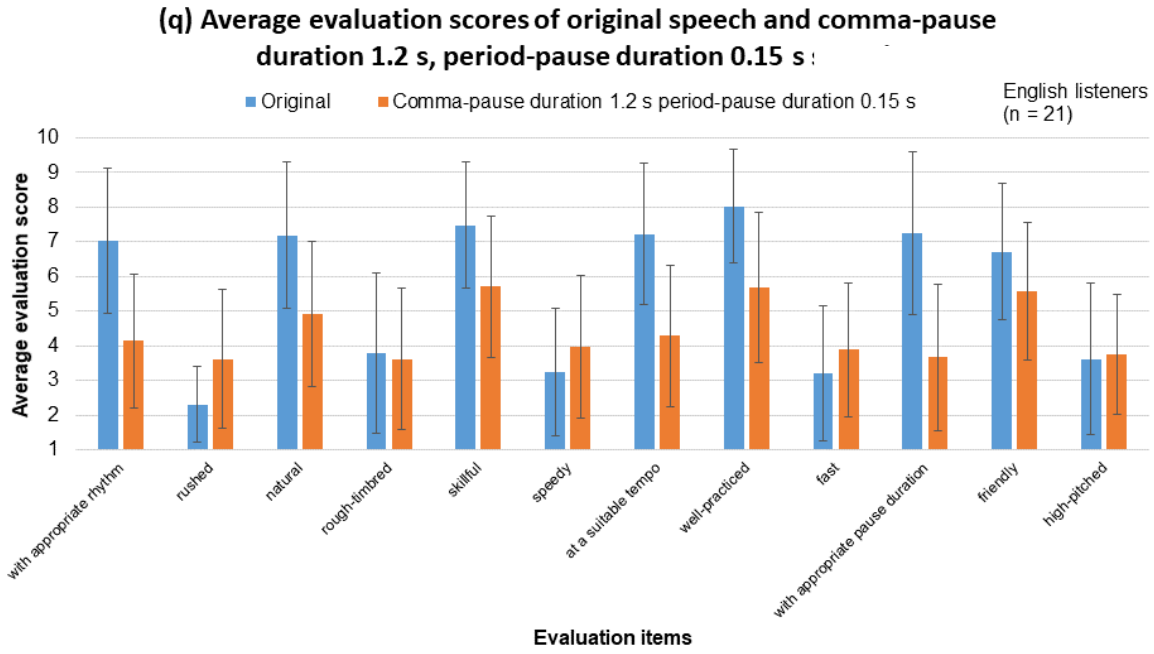

(n) Average evaluation scores of original speech and comma-pause duration 0.6 s, period-pause duration 0.6 s
Original
Comma-pause duration 0.6 s period-pause duration 0.6 s
English listeners (n = 21)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
with appropriate rhythm
rushed
natural
rough-timbred
skillful
speedy
at a suitable tempo
well-practiced
fast
with appropriate pause duration
friendly
high-pitched
Evaluation items

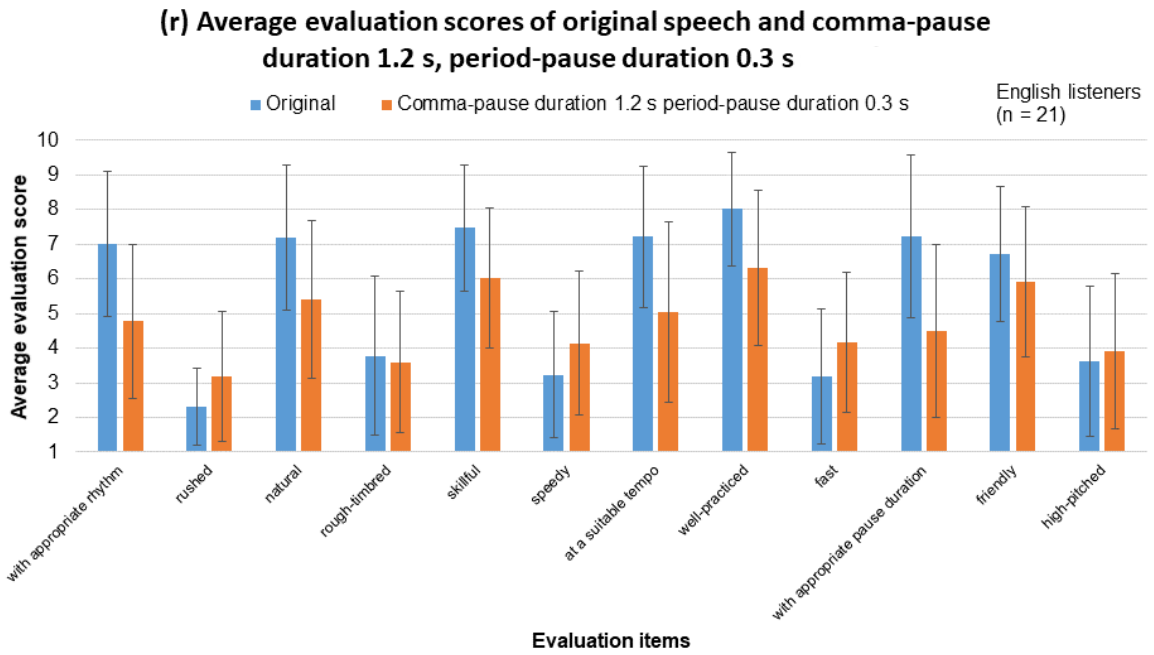
(o) Average evaluation scores of original speech and comma-pause duration 0.6 s, period-pause duration 1.2 s



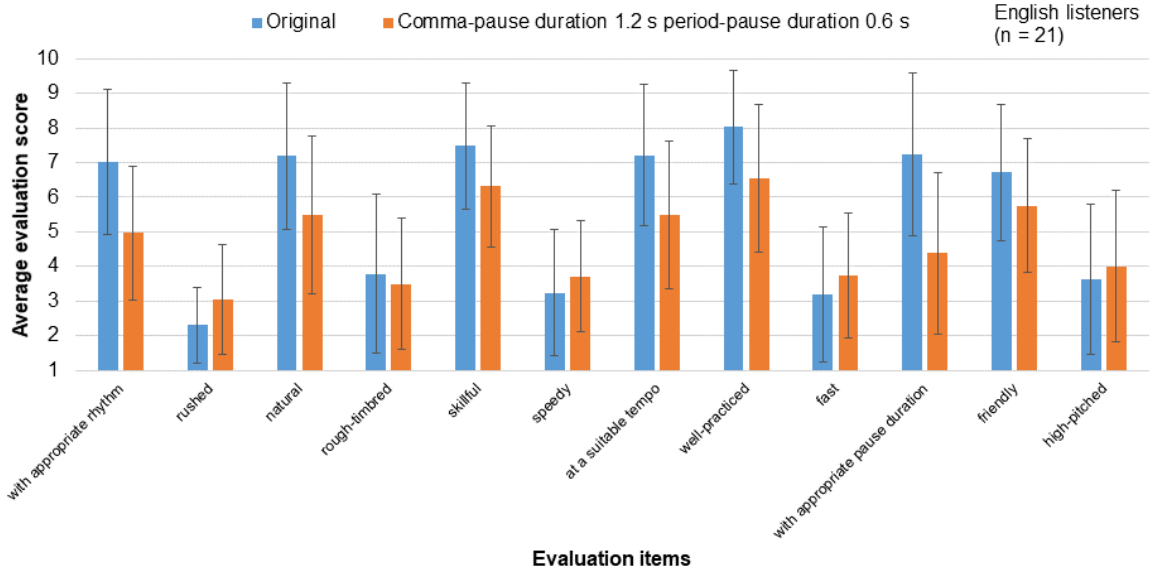
(p) Average evaluation scores of original speech and comma-pause duration 0.6 s, period-pause duration 2.4 s



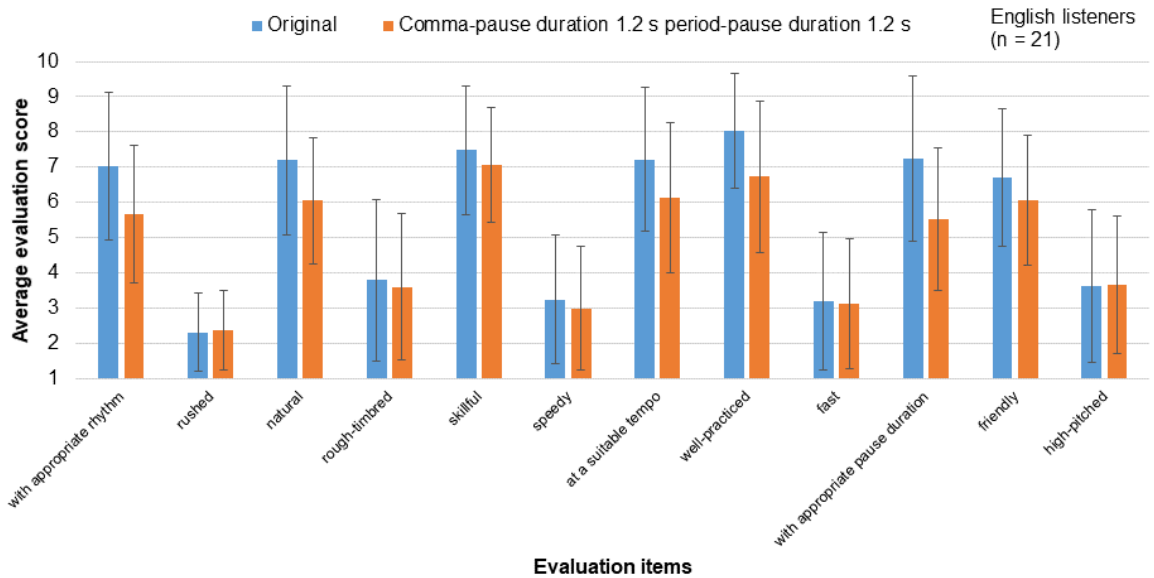

(q) Average evaluation scores of original speech and comma-pause duration 1.2 s, period-pause duration 0.15 s
Original
Comma-pause duration 1.2 s period-pause duration 0.15 s
English listeners (n = 21)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
with appropriate rhythm
rushed
natural
rough-timbred
skillful
speedy
at a suitable tempo
well-practiced
fast
with appropriate pause duration
friendly
high-pitched
Evaluation items


(r) Average evaluation scores of original speech and comma-pause duration 1.2 s, period-pause duration 0.3 s
Original
Comma-pause duration 1.2 s period-pause duration 0.3 s
English listeners (n = 21)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
with appropriate rhythm
rushed
natural
rough-timbred
skillful
speedy
at a suitable tempo
well-practiced
fast
with appropriate pause duration
friendly
high-pitched
Evaluation items

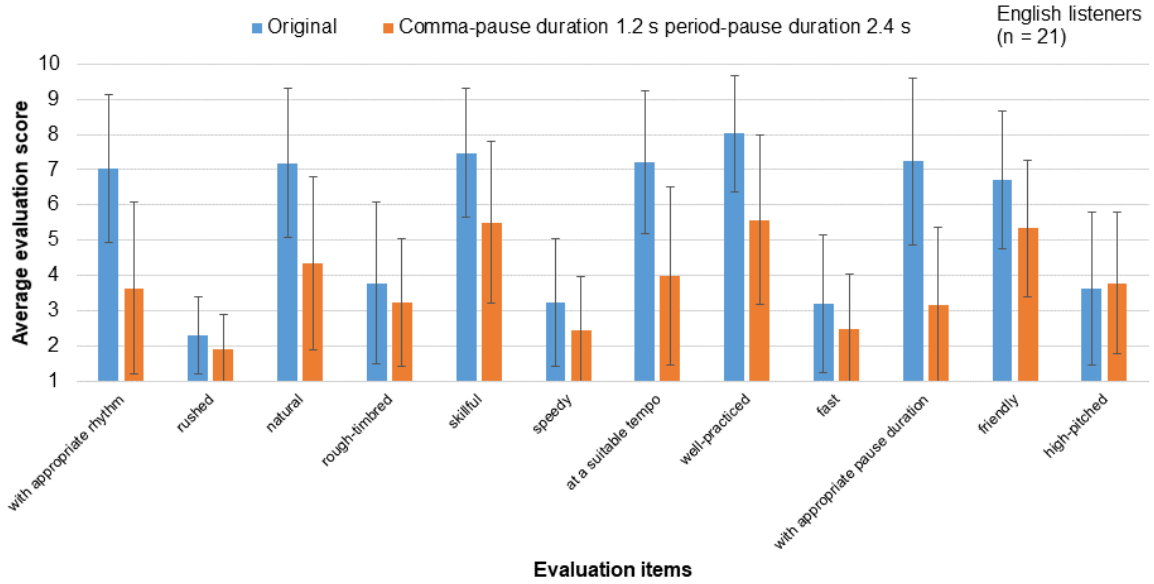
(s) Average evaluation scores of original speech and comma-pause duration 1.2 s, period-pause duration 0.6 s



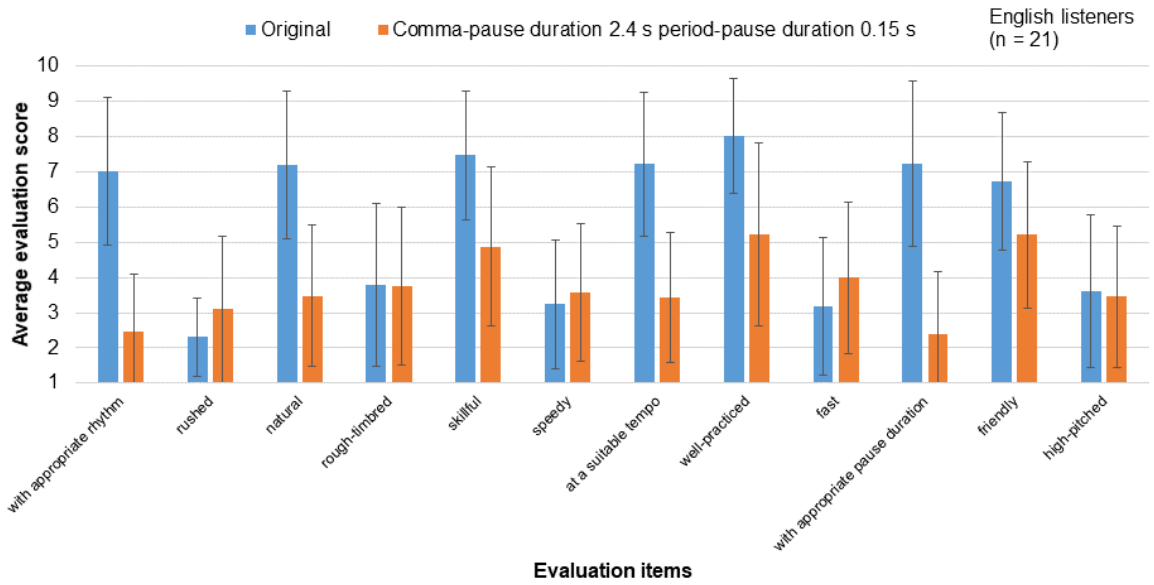
(t) Average evaluation scores of original speech and comma-pause duration 1.2 s, period-pause duration 1.2 s



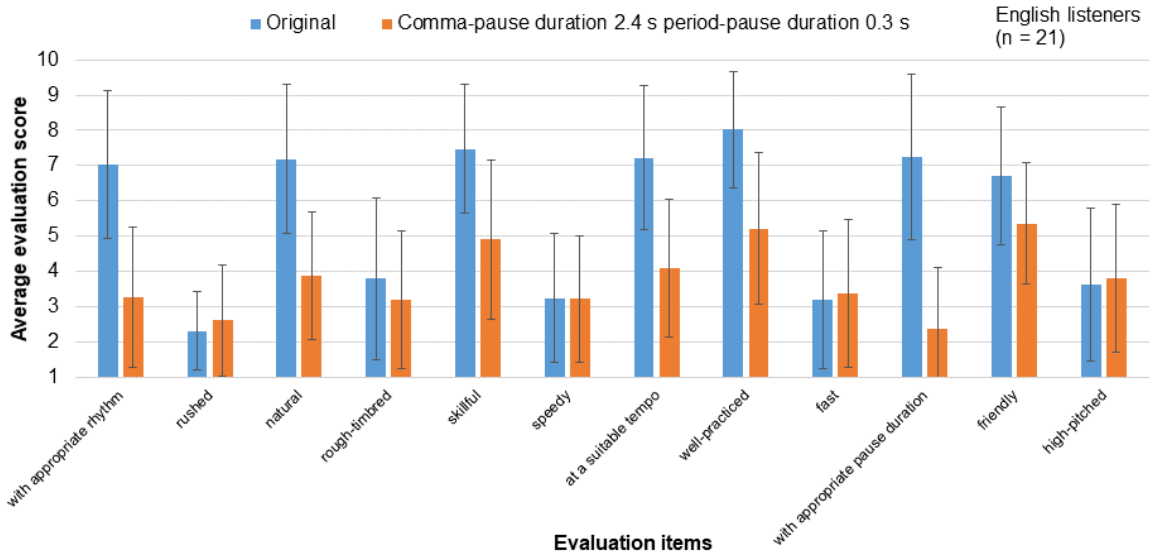
(u) Average evaluation scores of original speech and comma-pause duration 1.2 s, period-pause duration 2.4 s



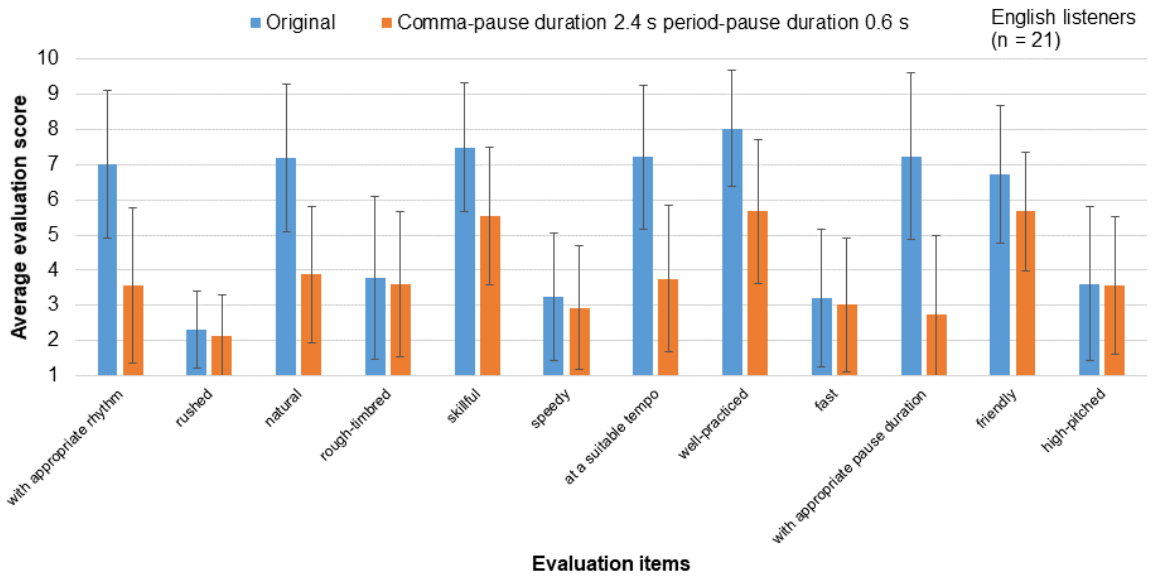
(v) Average evaluation scores of original speech and comma-pause duration 2.4 s, period-pause duration 0.15 s



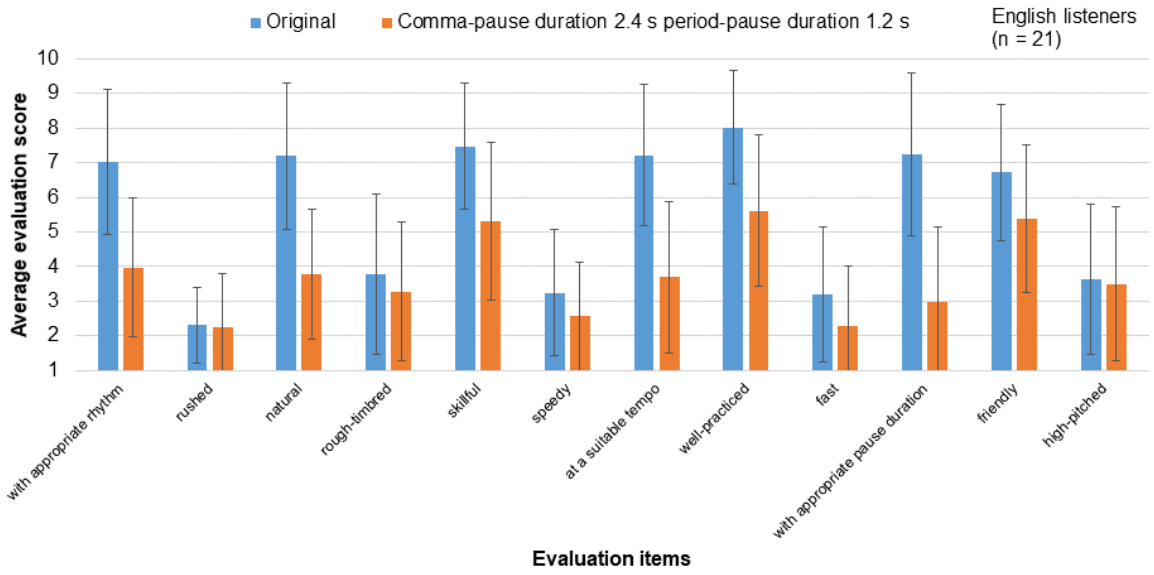
(w) Average evaluation scores of original speech and comma-pause duration 2.4 s, period-pause duration 0.3 s



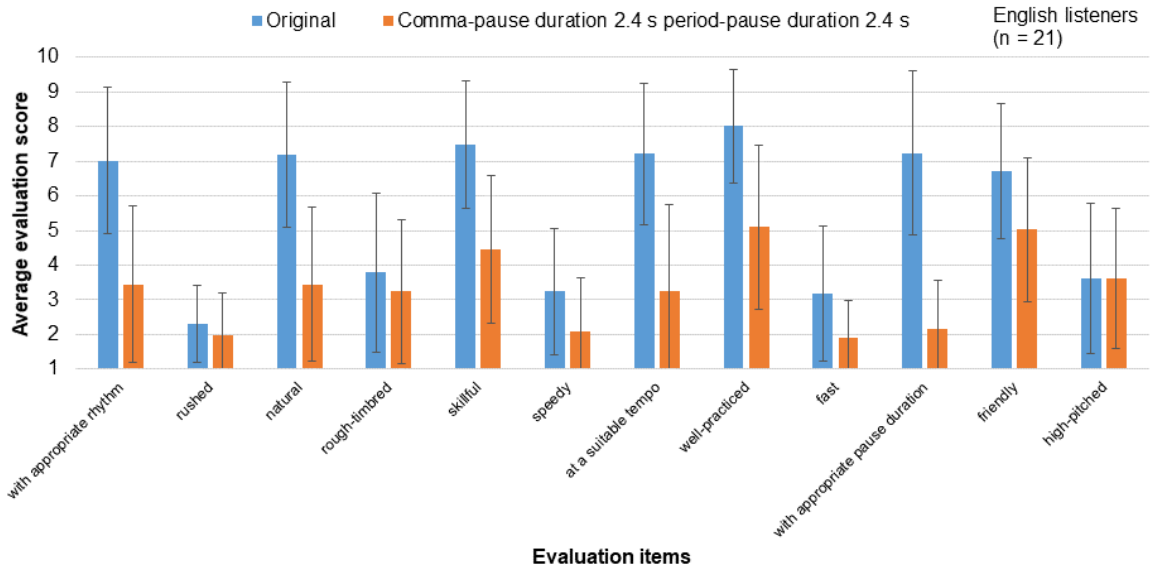
(x) Average evaluation scores of original speech and comma-pause duration 2.4 s, period-pause duration 0.6 s



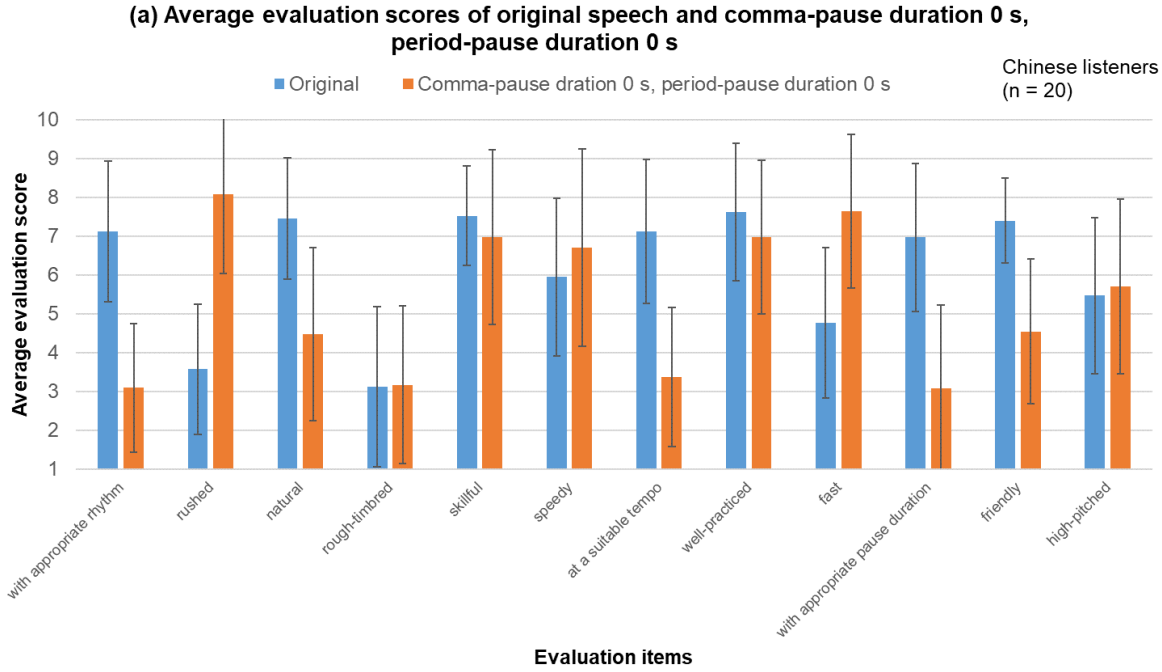
(y) Average evaluation scores of original speech and comma-pause duration 2.4 s, period-pause duration 1.2 s

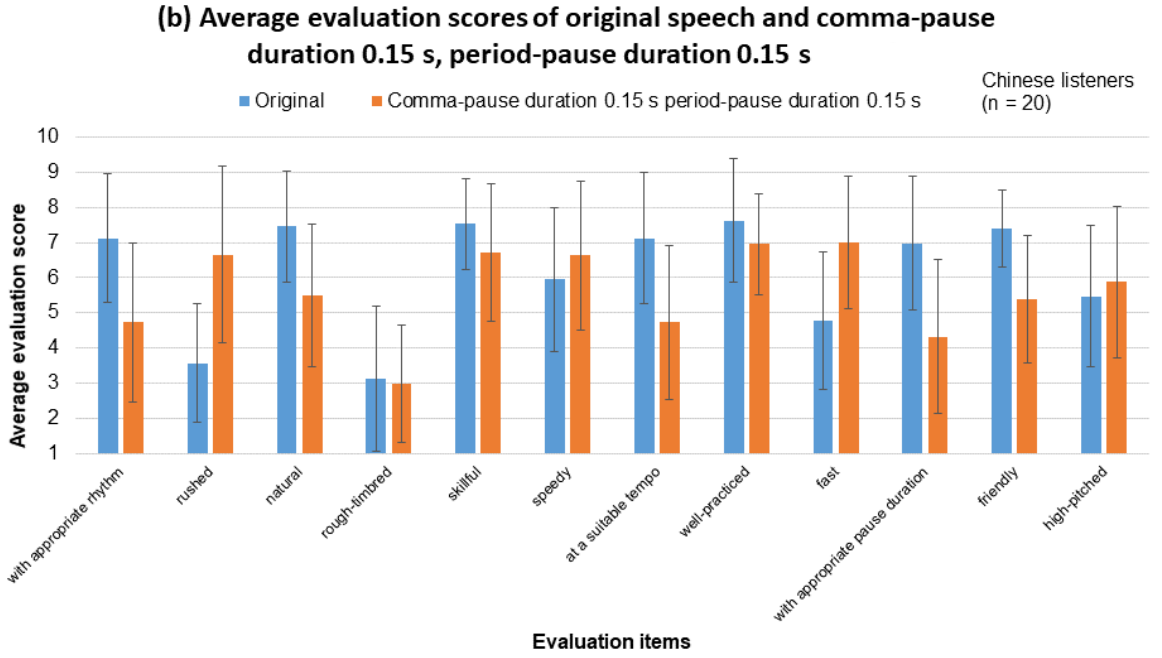


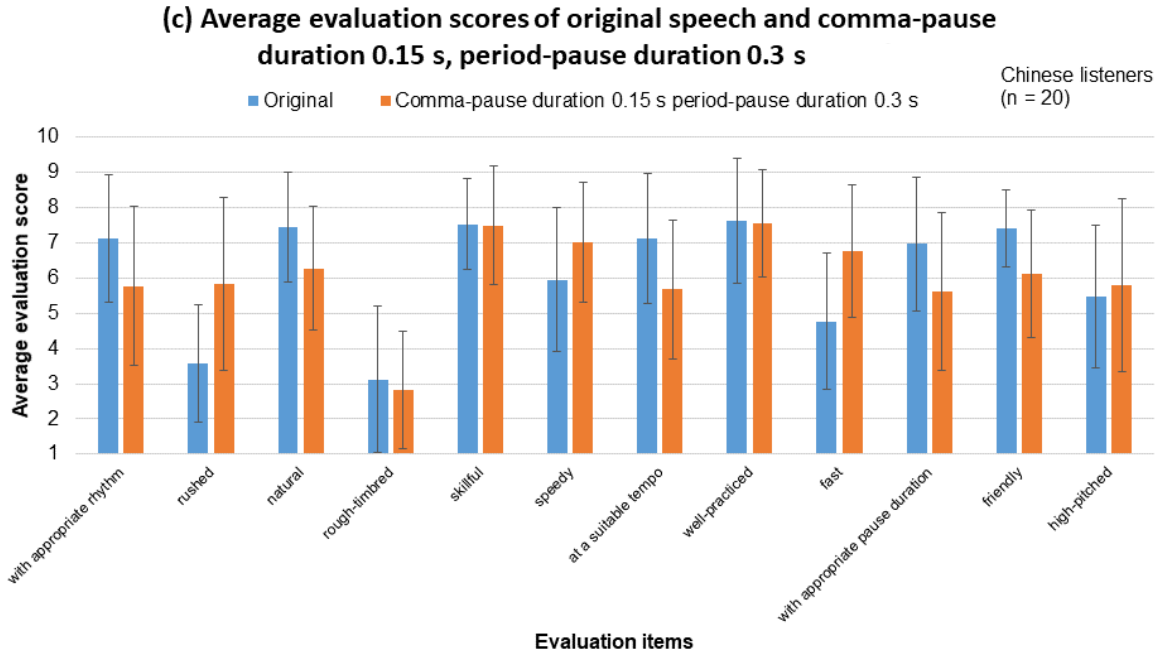
(z) Average evaluation scores of original speech and comma-pause duration 2.4 s, period-pause duration 2.4 s

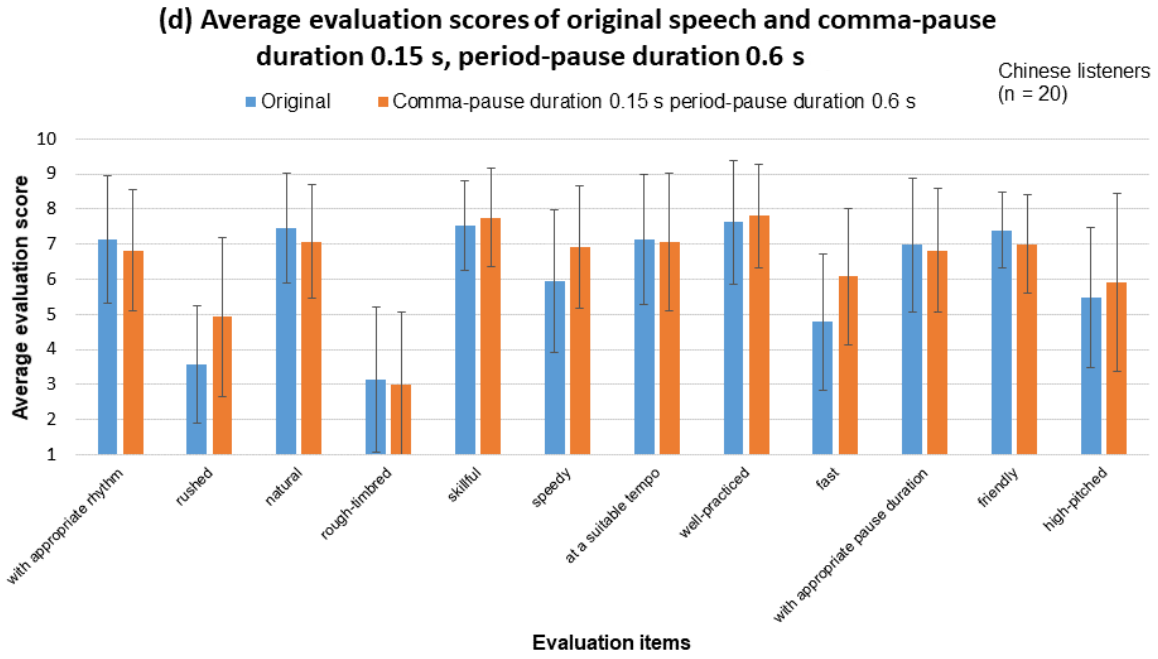


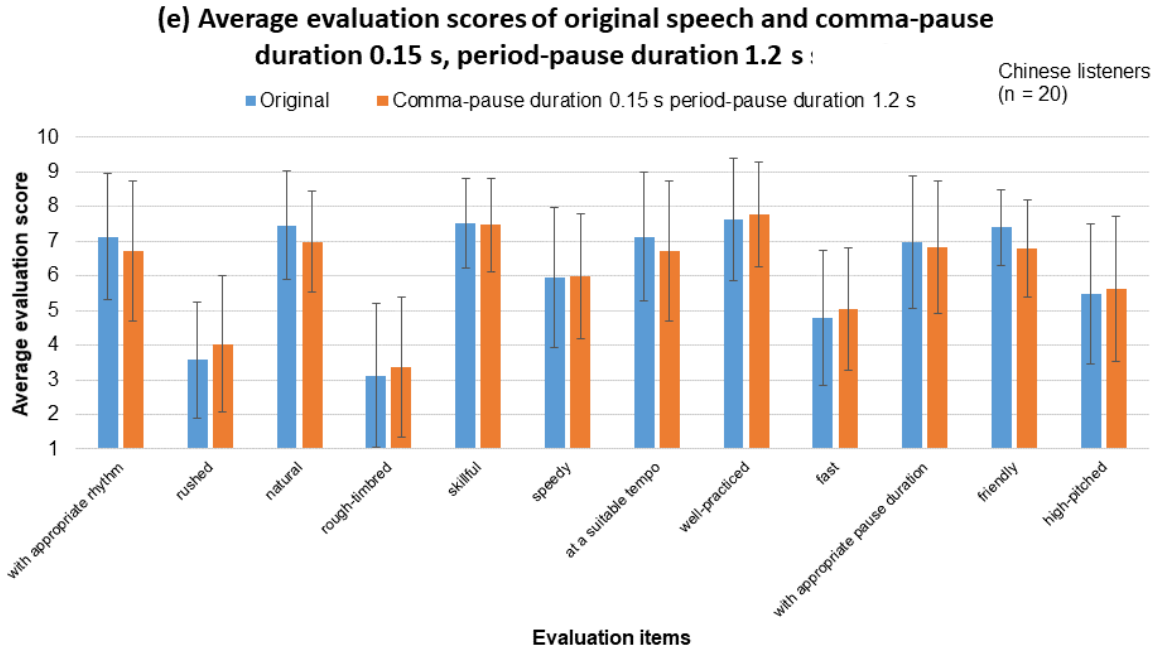
Chinese group

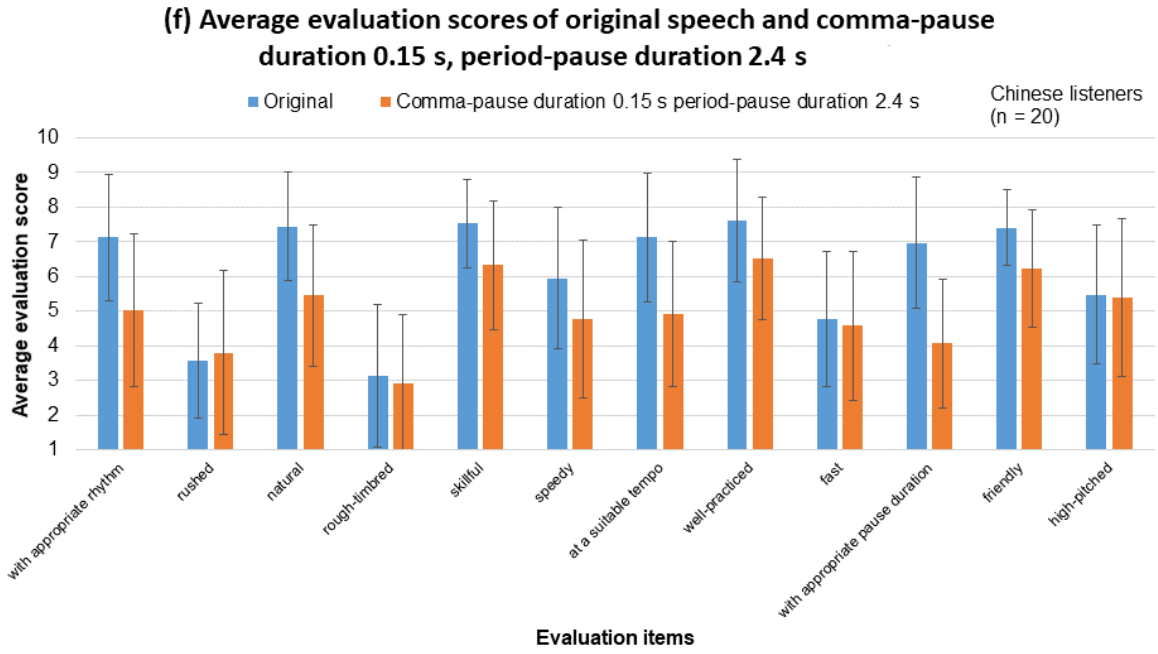


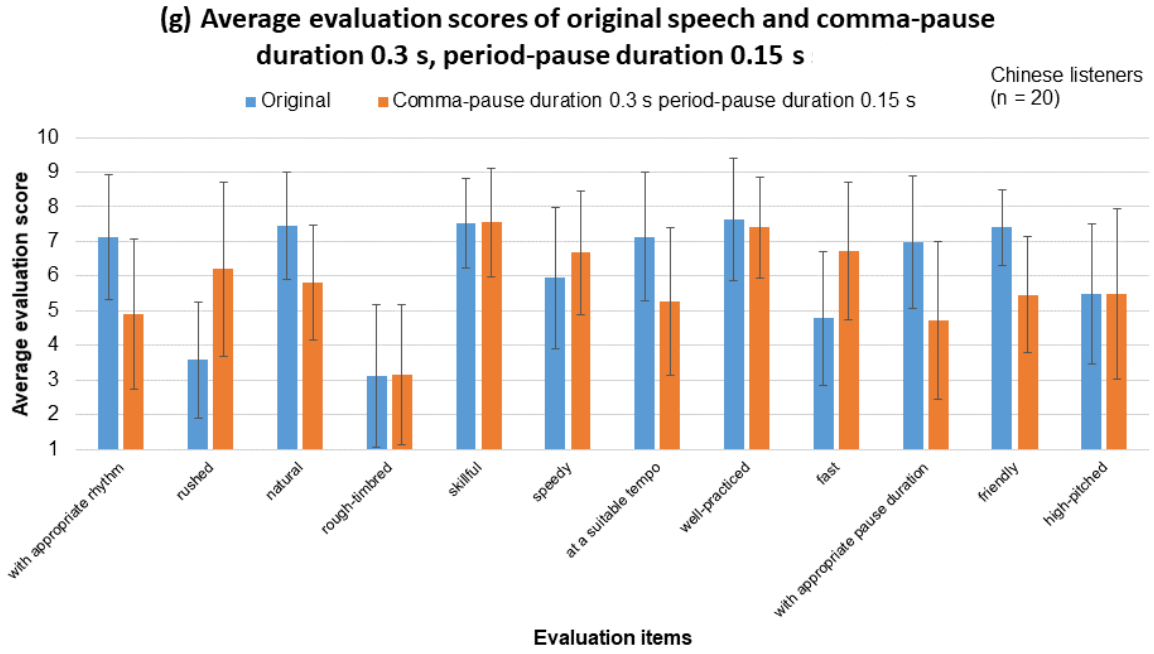


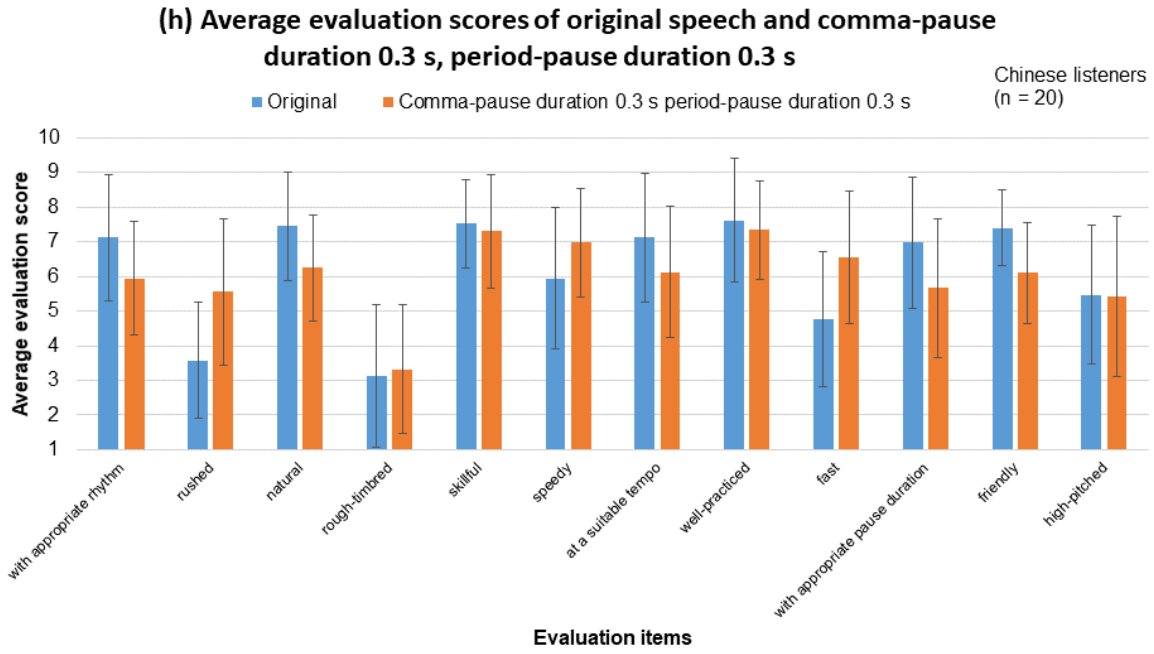

(c) Average evaluation scores of original speech and comma-pause duration 0.15 s, period-pause duration 0.3 s
Chinese listeners (n = 20)
Original
Comma-pause duration 0.15 s period-pause duration 0.3 s
Average evaluation score
10
9
8
7
6
5
4
3
2
1
with appropriate rhythm
rushed
natural
rough-timbred
skillful
speedy
at a suitable tempo
well-practiced
fast
with appropriate pause duration
friendly
high-pitched
Evaluation items

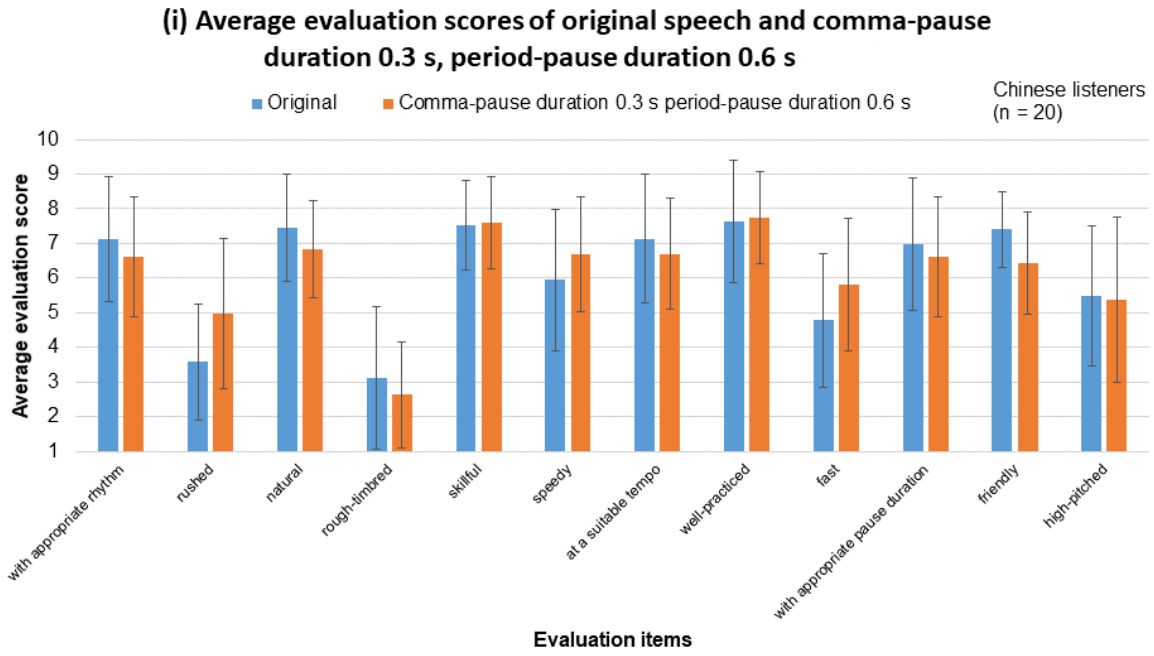

(d) Average evaluation scores of original speech and comma-pause duration 0.15 s, period-pause duration 0.6 s
Chinese listeners (n = 20)
Original
Comma-pause duration 0.15 s period-pause duration 0.6 s
Average evaluation score
10
9
8
7
6
5
4
3
2
1
with appropriate rhythm
rushed
natural
rough-timbred
skillful
speedy
at a suitable tempo
well-practiced
fast
with appropriate pause duration
friendly
high-pitched
Evaluation items

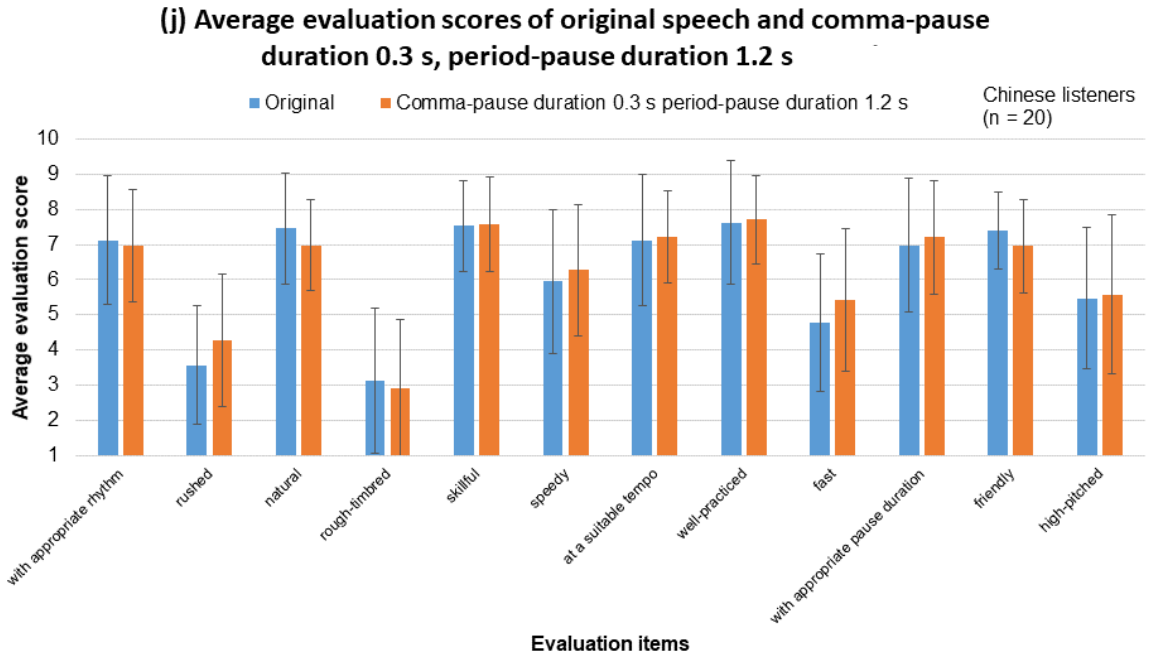

(e) Average evaluation scores of original speech and comma-pause duration 0.15 s, period-pause duration 1.2 s
Chinese listeners (n = 20)
Original
Comma-pause duration 0.15 s period-pause duration 1.2 s
Average evaluation score
10
9
8
7
6
5
4
3
2
1
with appropriate rhythm
rushed
natural
rough-timbred
skillful
speedy
at a suitable tempo
well-practiced
fast
with appropriate pause duration
friendly
high-pitched
Evaluation items


(f) Average evaluation scores of original speech and comma-pause duration 0.15 s, period-pause duration 2.4 s
Chinese listeners (n = 20)
Original
Comma-pause duration 0.15 s period-pause duration 2.4 s
Average evaluation score
10
9
8
7
6
5
4
3
2
1
with appropriate rhythm
rushed
natural
rough-timbred
skillful
speedy
at a suitable tempo
well-practiced
fast
with appropriate pause duration
friendly
high-pitched
Evaluation items

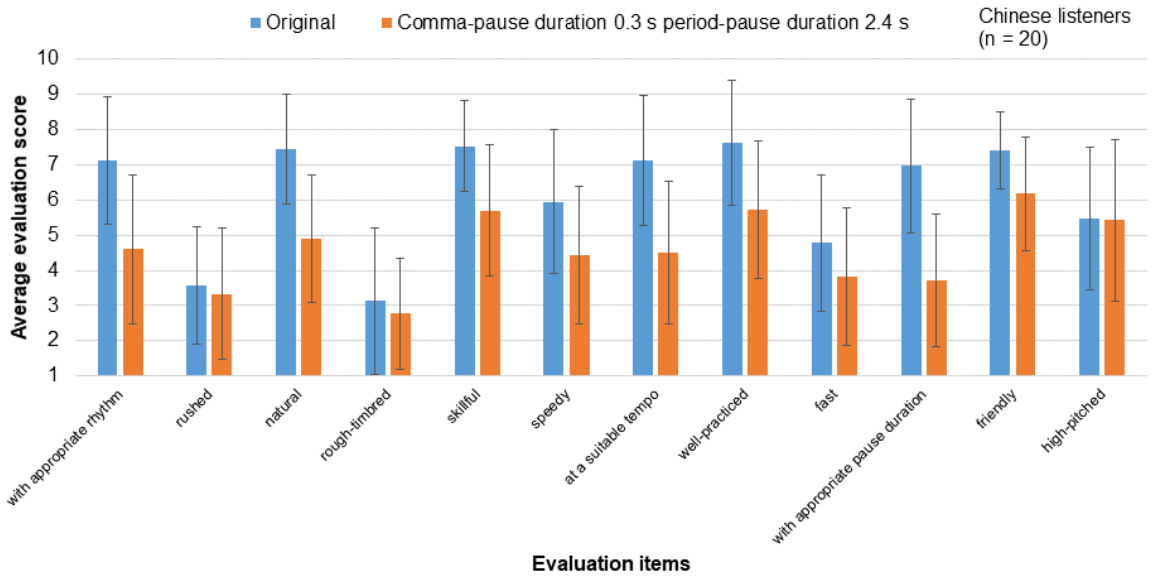

(g) Average evaluation scores of original speech and comma-pause duration 0.3 s, period-pause duration 0.15 s
Chinese listeners (n = 20)
Original
Comma-pause duration 0.3 s period-pause duration 0.15 s
Average evaluation score
10
9
8
7
6
5
4
3
2
1
with appropriate rhythm
rushed
natural
rough-timbred
skillful
speedy
at a suitable tempo
well-practiced
fast
with appropriate pause duration
friendly
high-pitched
Evaluation items


(h) Average evaluation scores of original speech and comma-pause duration 0.3 s, period-pause duration 0.3 s
Chinese listeners (n = 20)
Original
Comma-pause duration 0.3 s period-pause duration 0.3 s
Average evaluation score
10
9
8
7
6
5
4
3
2
1
with appropriate rhythm
rushed
natural
rough-timbred
skillful
speedy
at a suitable tempo
well-practiced
fast
with appropriate pause duration
friendly
high-pitched
Evaluation items

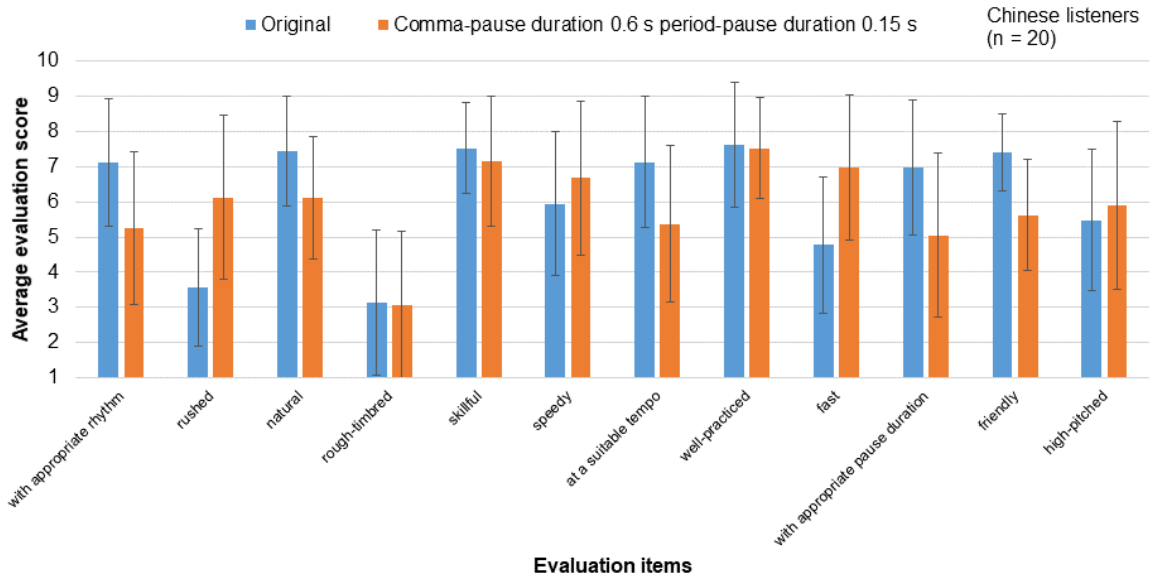

(i) Average evaluation scores of original speech and comma-pause duration 0.3 s, period-pause duration 0.6 s
Original
Comma-pause duration 0.3 s period-pause duration 0.6 s
Chinese listeners (n = 20)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
with appropriate rhythm
rushed
natural
rough-timbred
skillful
speedy
at a suitable tempo
well-practiced
fast
with appropriate pause duration
friendly
high-pitched
Evaluation items

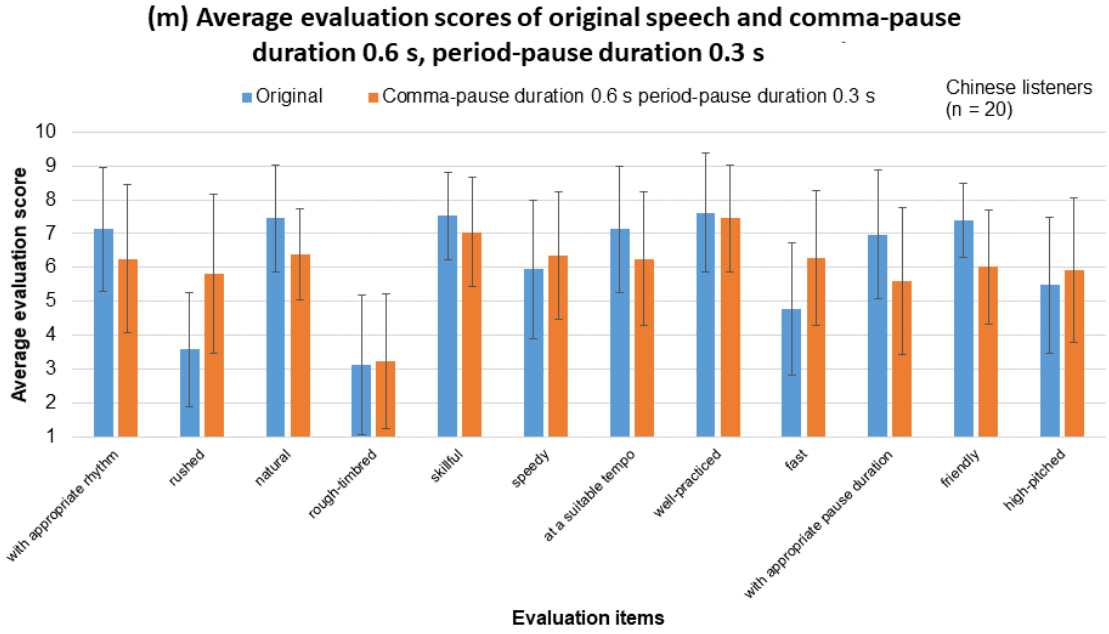

(j) Average evaluation scores of original speech and comma-pause duration 0.3 s, period-pause duration 1.2 s
Original
Comma-pause duration 0.3 s period-pause duration 1.2 s
Chinese listeners (n = 20)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
with appropriate rhythm
rushed
natural
rough-timbred
skillful
speedy
at a suitable tempo
well-practiced
fast
with appropriate pause duration
friendly
high-pitched
Evaluation items

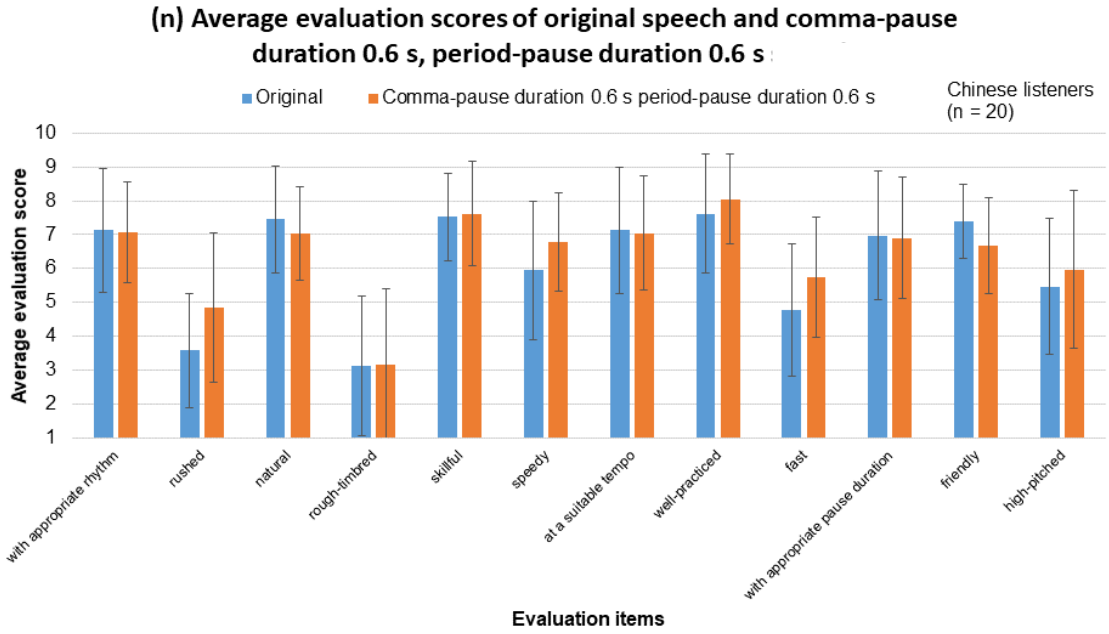
(k) Average evaluation scores of original speech and comma-pause duration 0.3 s, period-pause duration 2.4 s

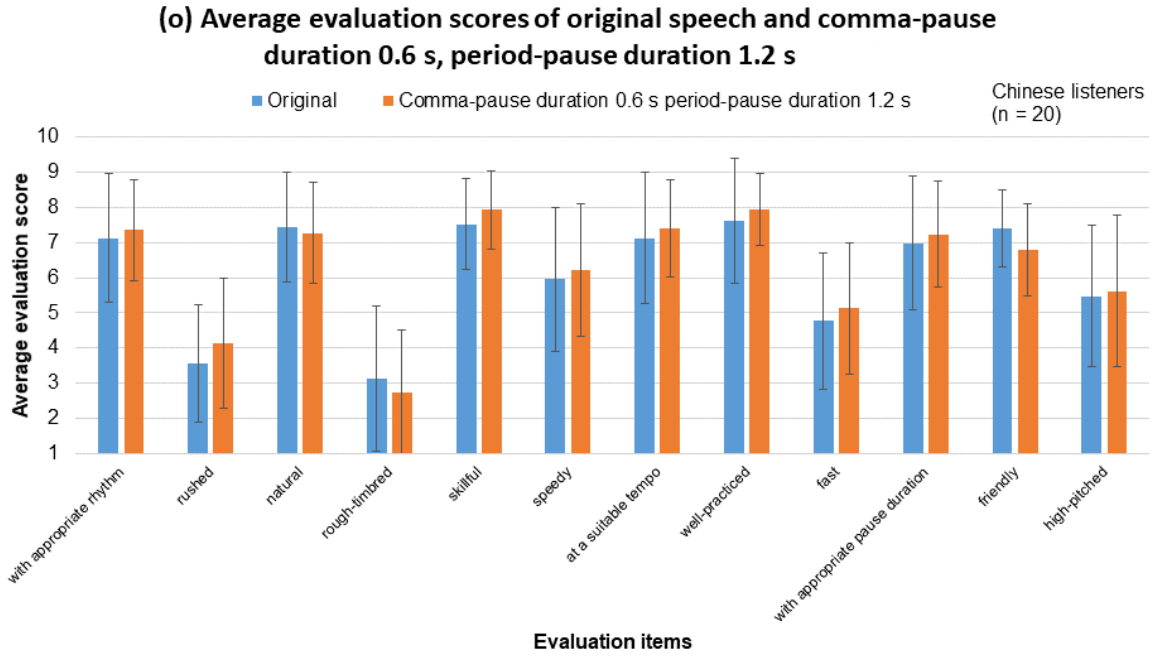


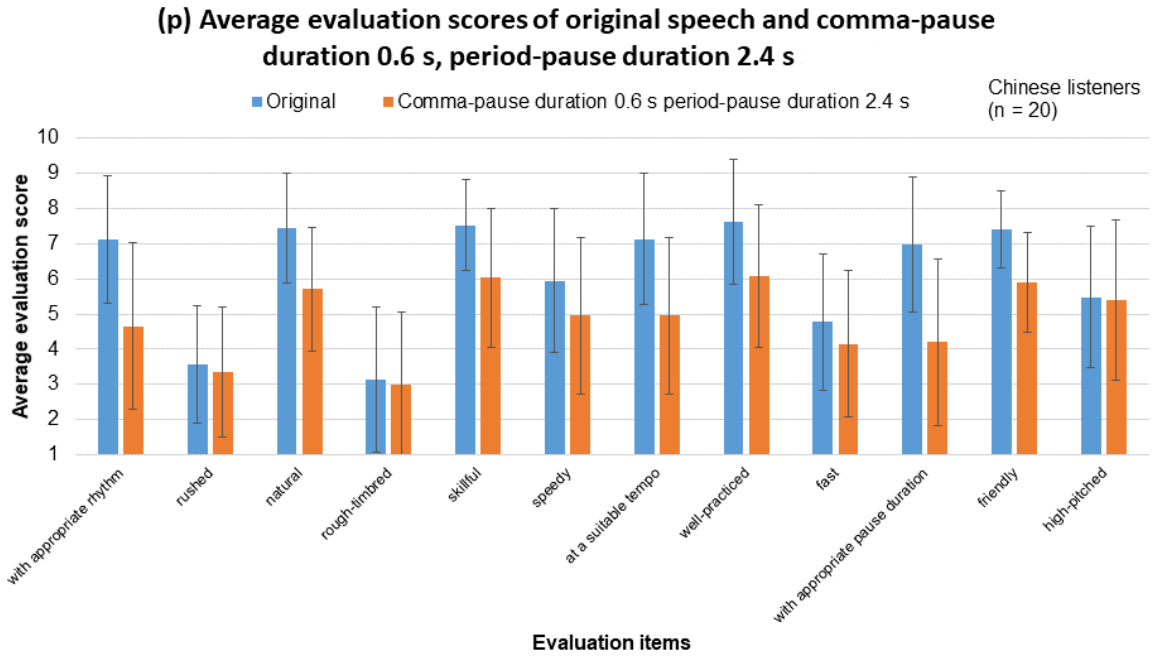
(l) Average evaluation scores of original speech and comma-pause duration 0.6 s, period-pause duration 0.15 s



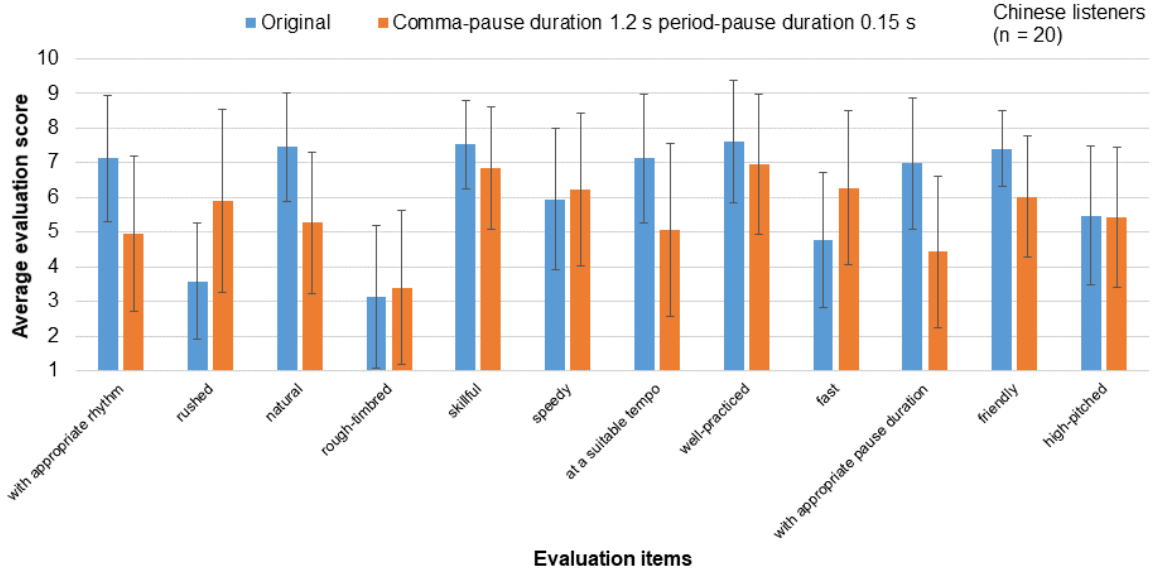

(m) Average evaluation scores of original speech and comma-pause duration 0.6 s, period-pause duration 0.3 s
Original
Comma-pause duration 0.6 s period-pause duration 0.3 s
Chinese listeners (n = 20)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
with appropriate rhythm
rushed
natural
rough-timbred
skillful
speedy
at a suitable tempo
well-practiced
fast
with appropriate pause duration
friendly
high-pitched
Evaluation items


(n) Average evaluation scores of original speech and comma-pause duration 0.6 s, period-pause duration 0.6 s
Original
Comma-pause duration 0.6 s period-pause duration 0.6 s
Chinese listeners (n = 20)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
with appropriate rhythm
rushed
natural
rough-timbred
skillful
speedy
at a suitable tempo
well-practiced
fast
with appropriate pause duration
friendly
high-pitched
Evaluation items

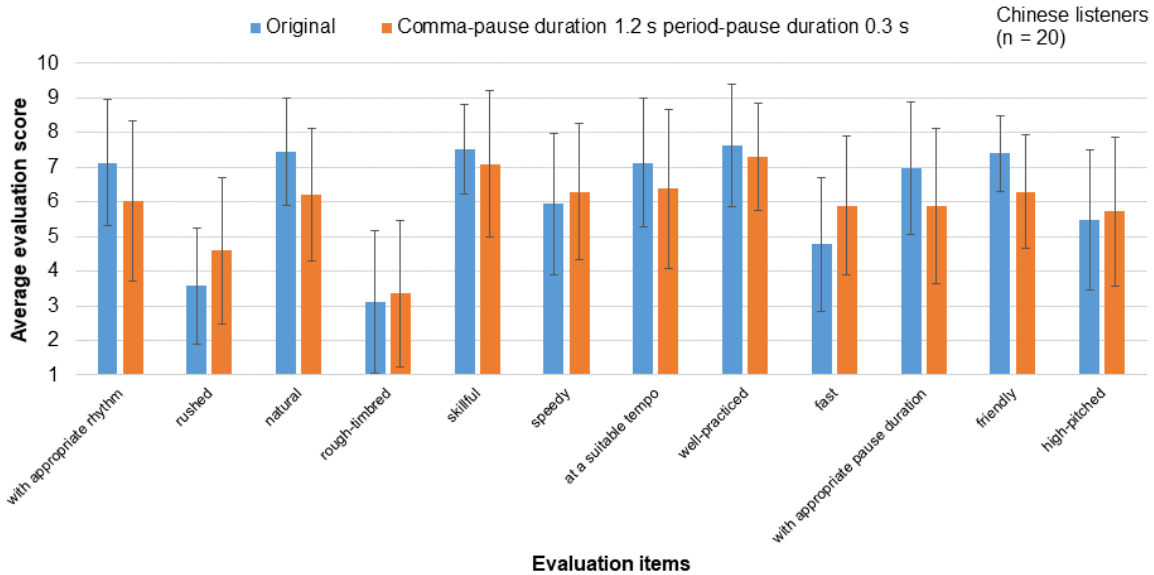

(o) Average evaluation scores of original speech and comma-pause duration 0.6 s, period-pause duration 1.2 s
Original
Comma-pause duration 0.6 s period-pause duration 1.2 s
Chinese listeners (n = 20)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
with appropriate rhythm
rushed
natural
rough-timbred
skillful
speedy
at a suitable tempo
well-practiced
fast
with appropriate pause duration
friendly
high-pitched
Evaluation items


(p) Average evaluation scores of original speech and comma-pause duration 0.6 s, period-pause duration 2.4 s
Original
Comma-pause duration 0.6 s period-pause duration 2.4 s
Chinese listeners (n = 20)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
with appropriate rhythm
rushed
natural
rough-timbred
skillful
speedy
at a suitable tempo
well-practiced
fast
with appropriate pause duration
friendly
high-pitched
Evaluation items

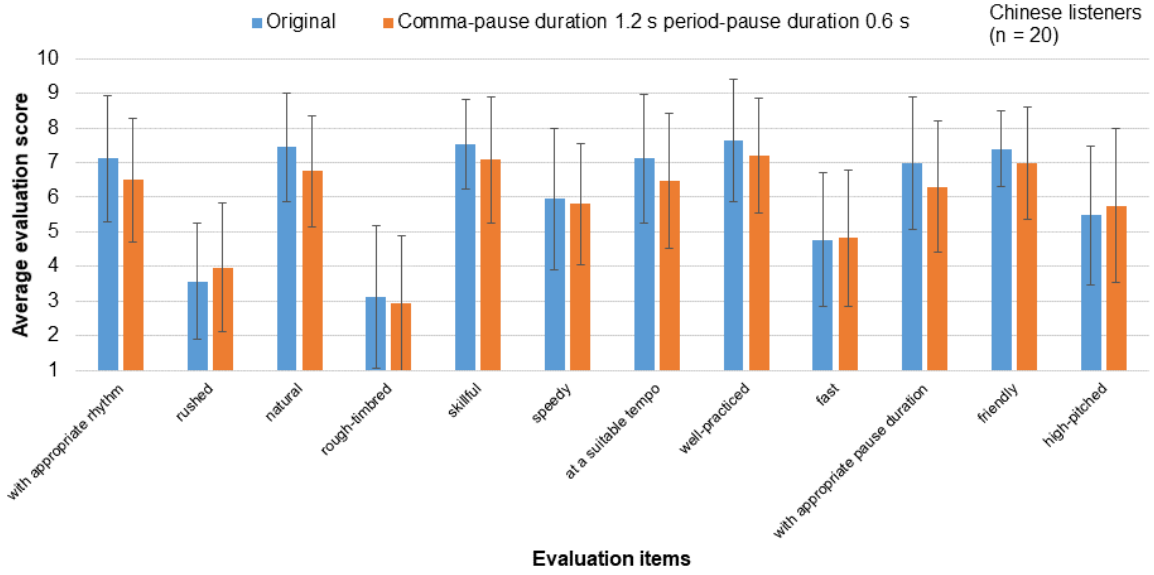
(q) Average evaluation scores of original speech and comma-pause duration 1.2 s, period-pause duration 0.15 s



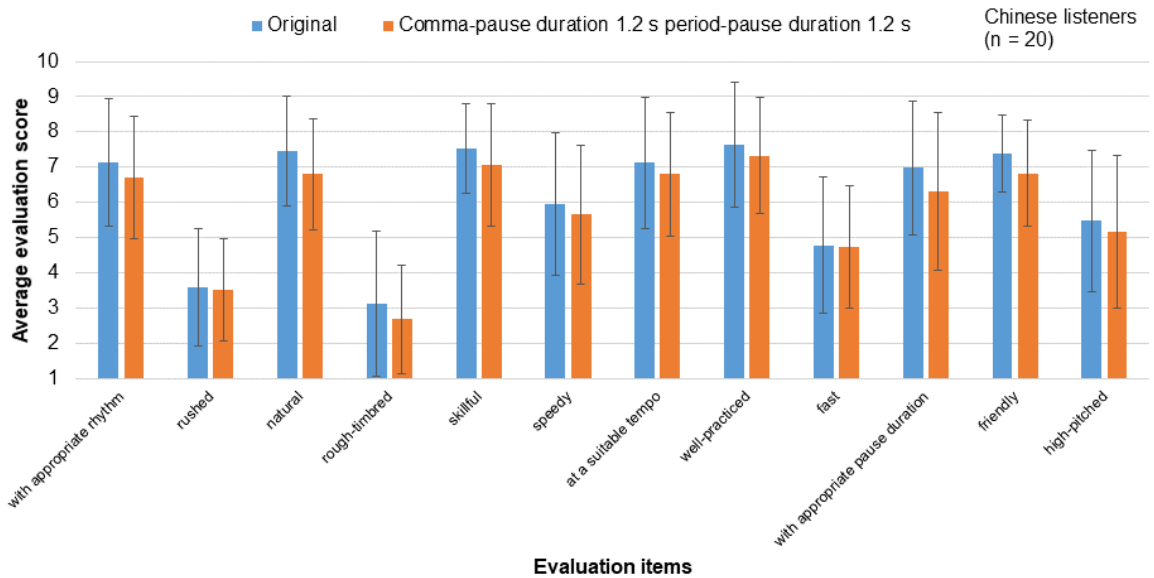
(r) Average evaluation scores of original speech and comma-pause duration 1.2 s, period-pause duration 0.3 s



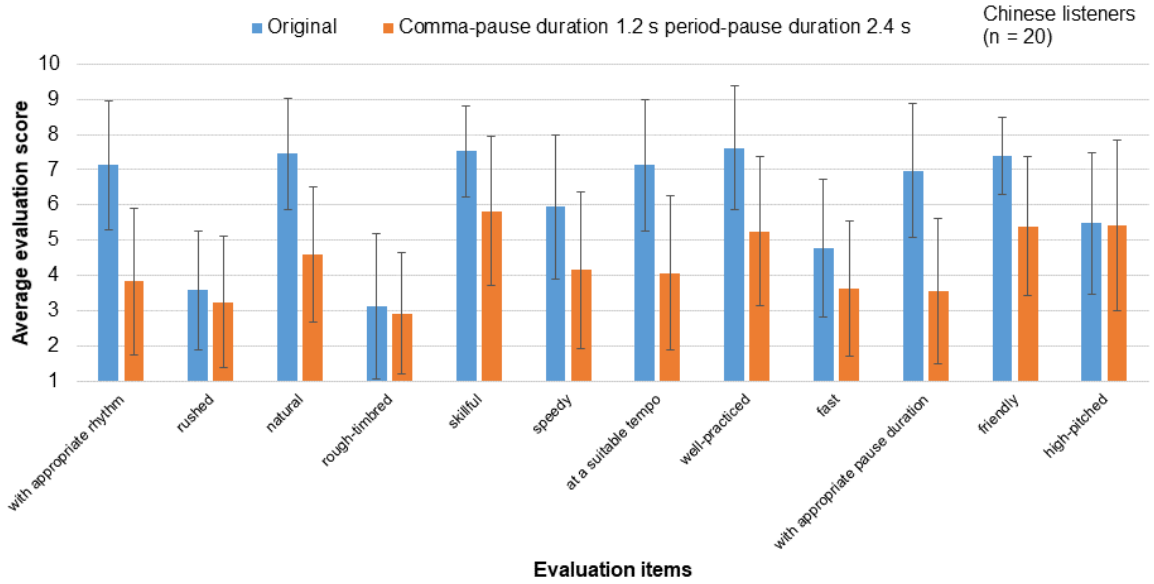
(s) Average evaluation scores of original speech and comma-pause duration 1.2 s, period-pause duration 0.6 s



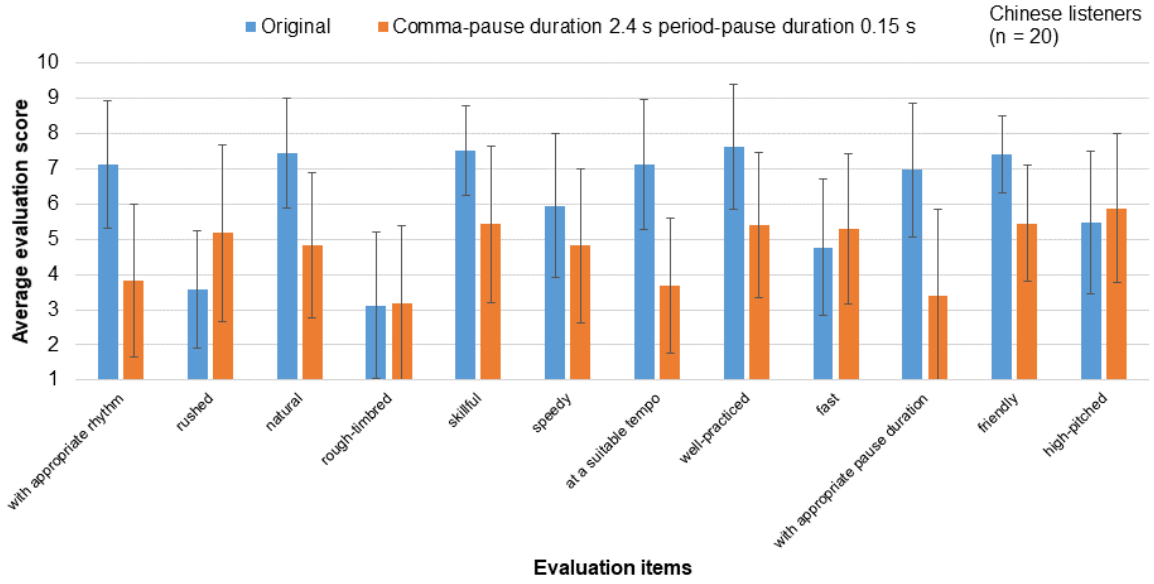
(t) Average evaluation scores of original speech and comma-pause duration 1.2 s, period-pause duration 1.2 s



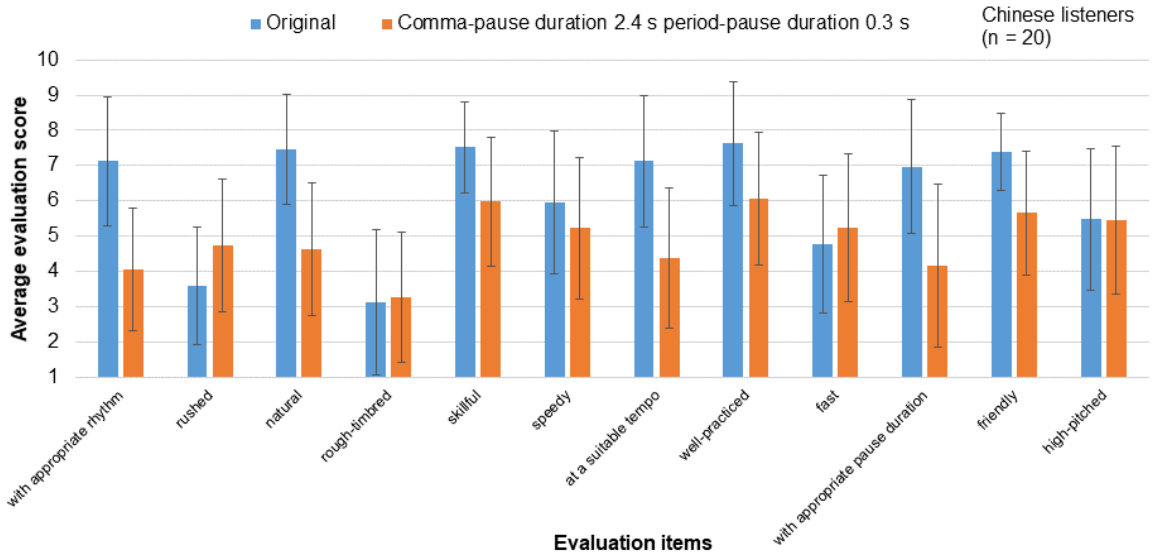
(u) Average evaluation scores of original speech and comma-pause duration 1.2 s, period-pause duration 2.4 s



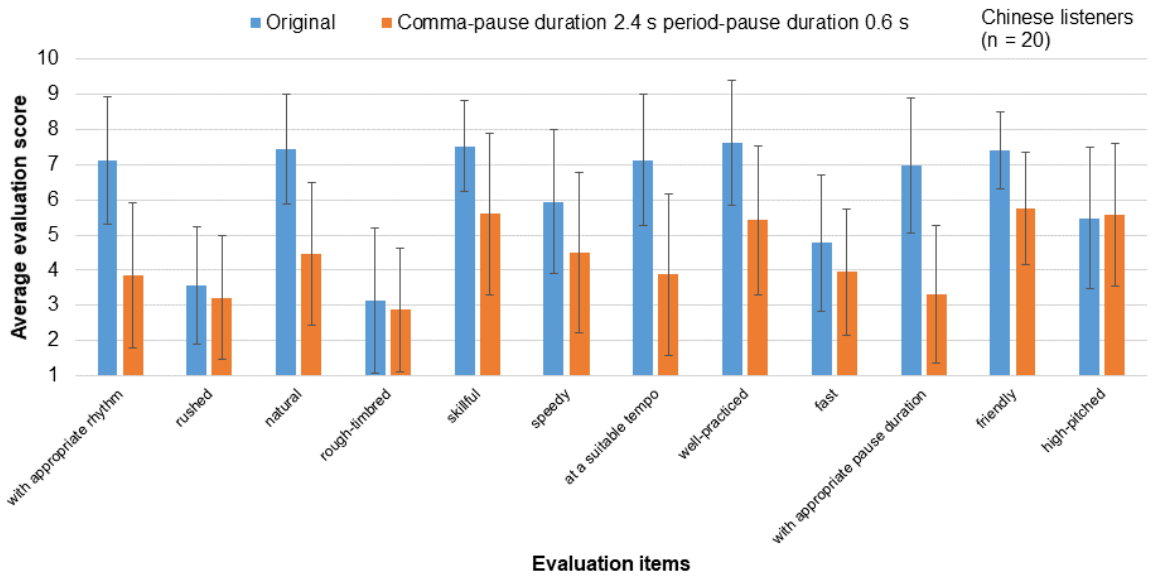
(v) Average evaluation scores of original speech and comma-pause duration 2.4 s, period-pause duration 0.15 s



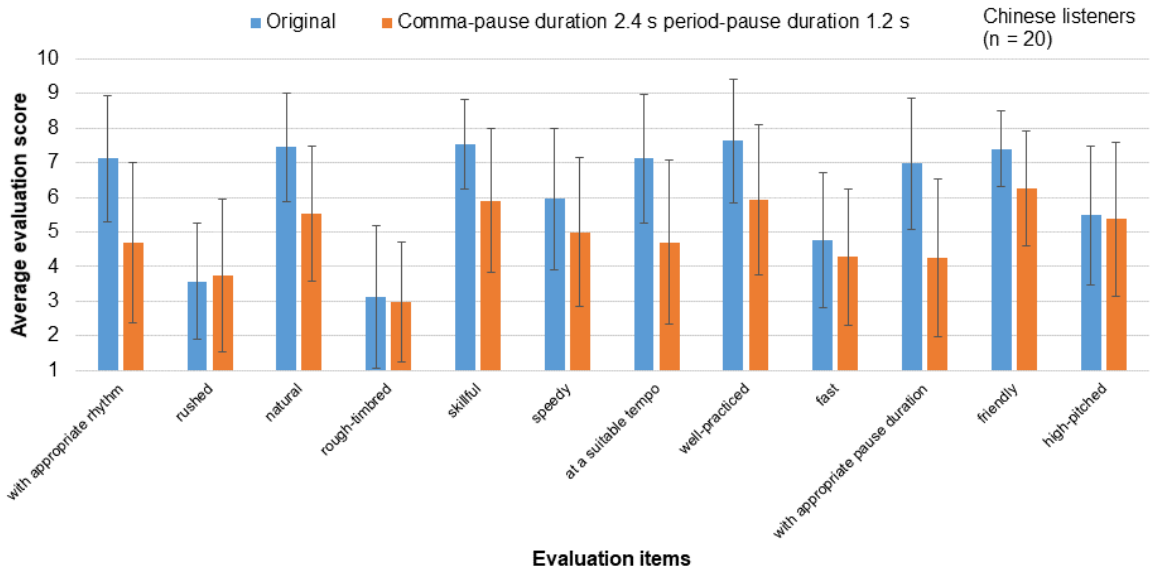
(w) Average evaluation scores of original speech and comma-pause duration 2.4 s, period-pause duration 0.3 s



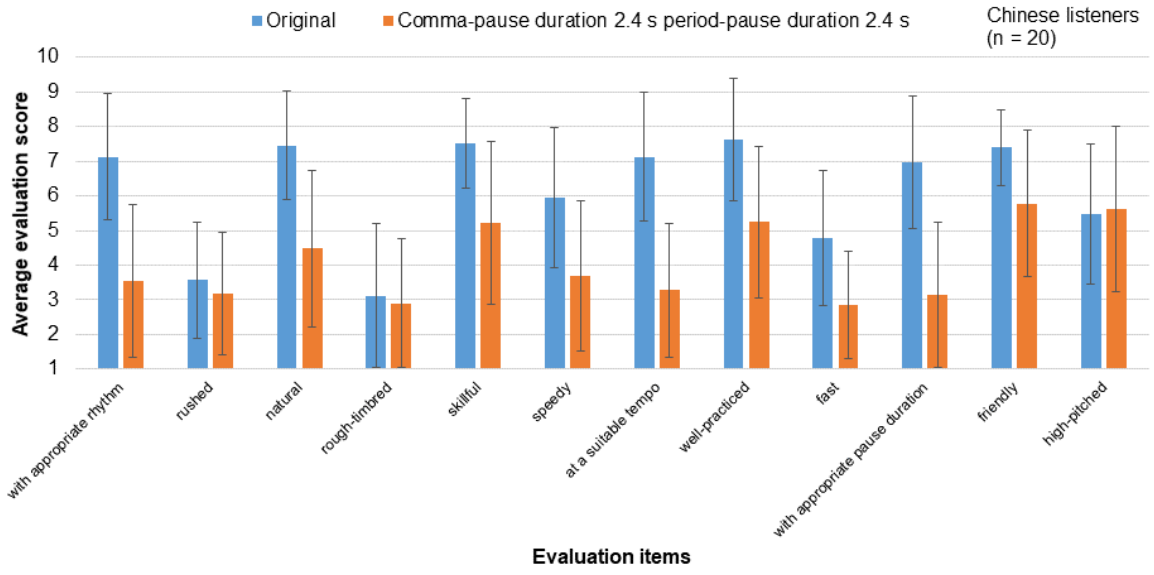
(x) Average evaluation scores of original speech and comma-pause duration 2.4 s, period-pause duration 0.6 s



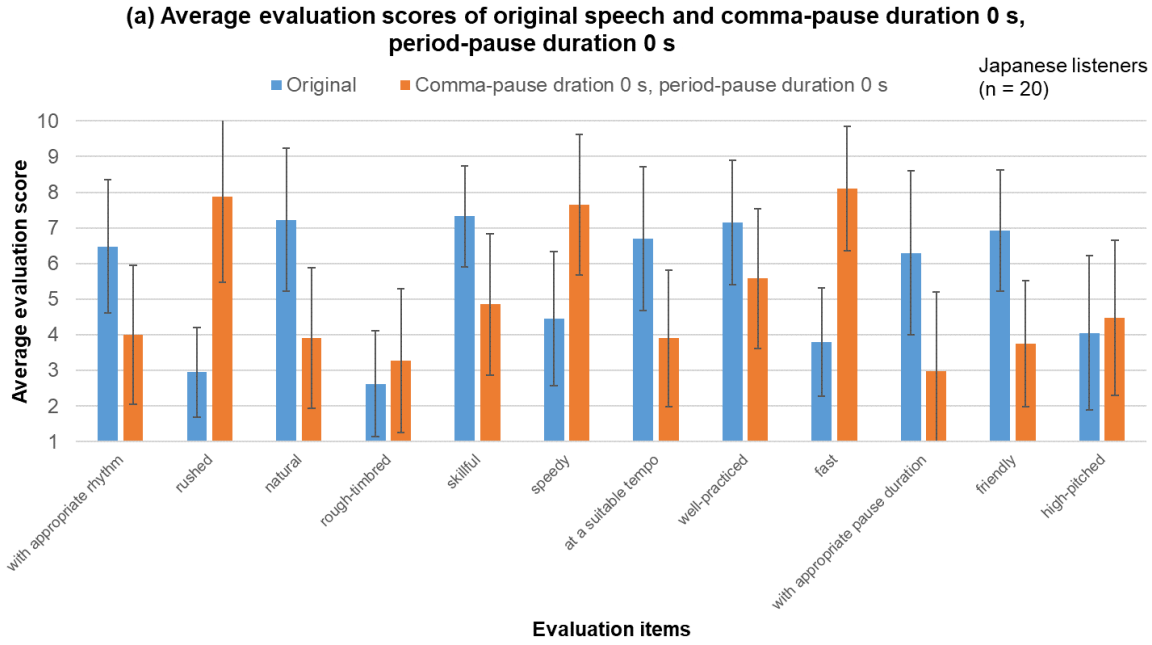
(y) Average evaluation scores of original speech and comma-pause duration 2.4 s, period-pause duration 1.2 s

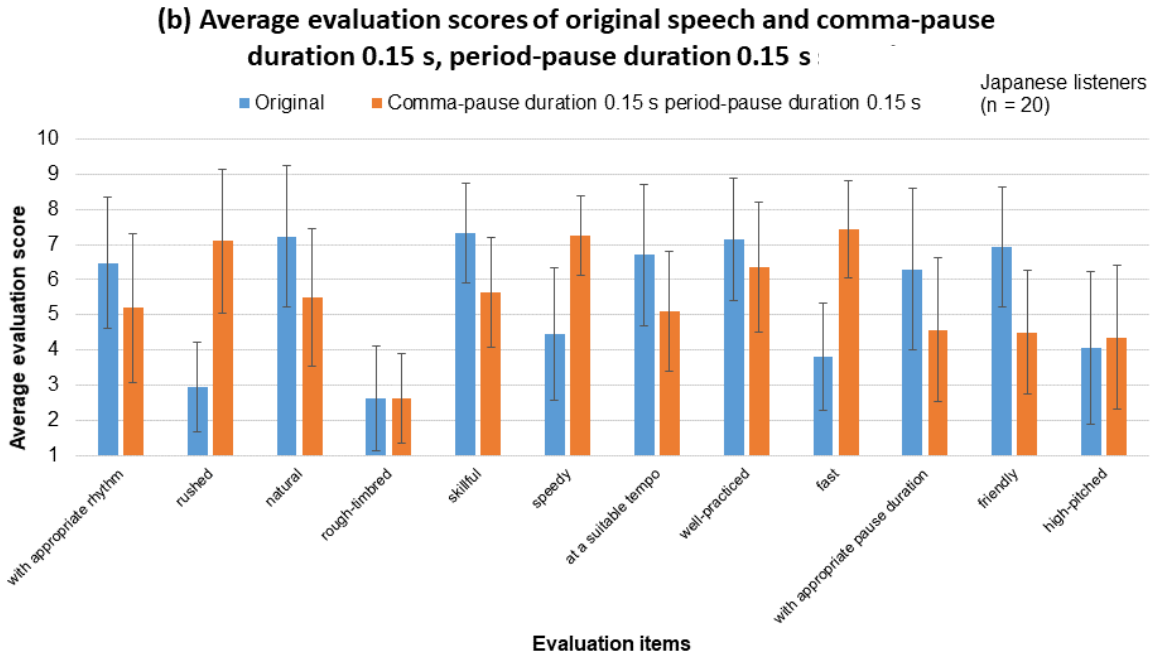


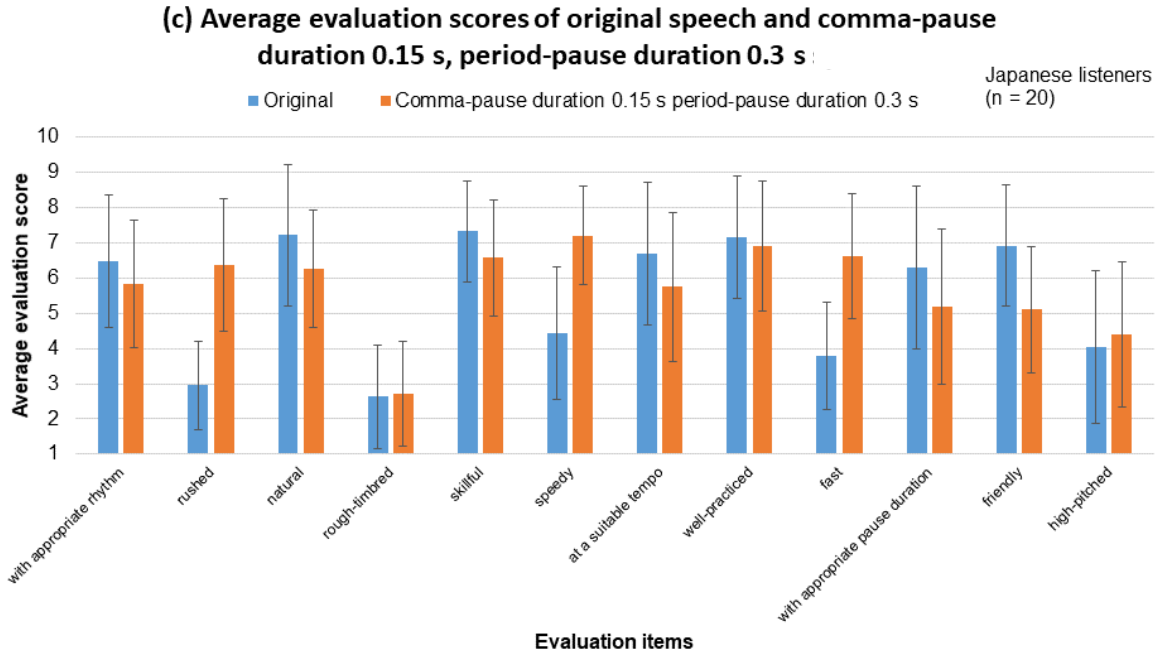
(z) Average evaluation scores of original speech and comma-pause duration 2.4 s, period-pause duration 2.4 s

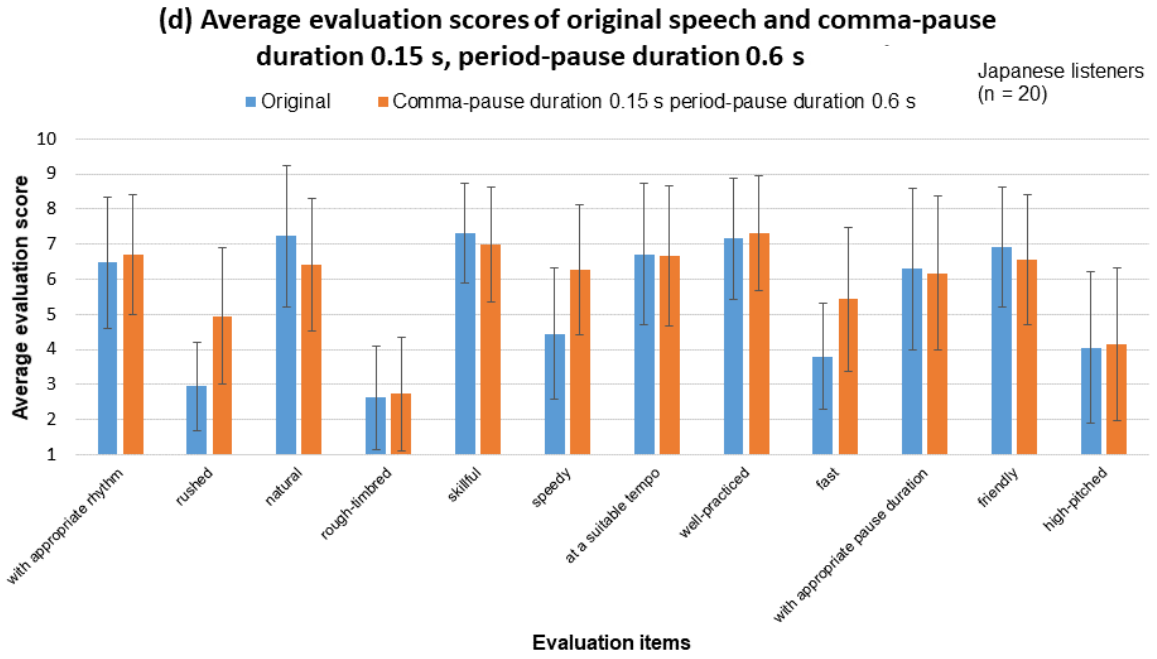


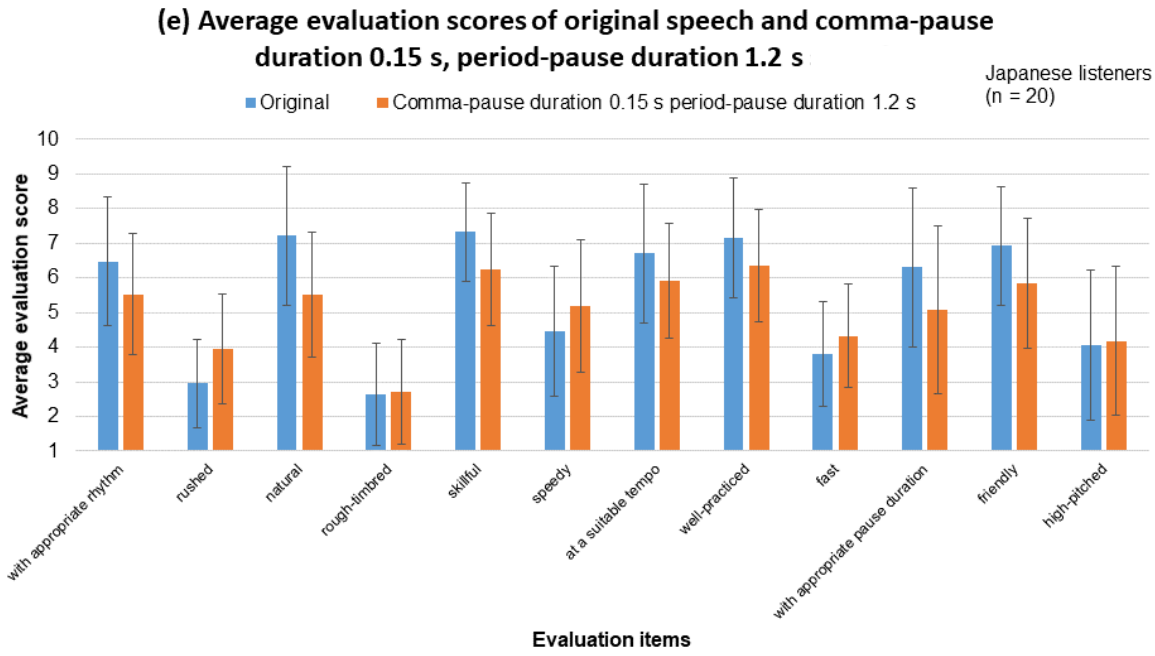
Japanese group

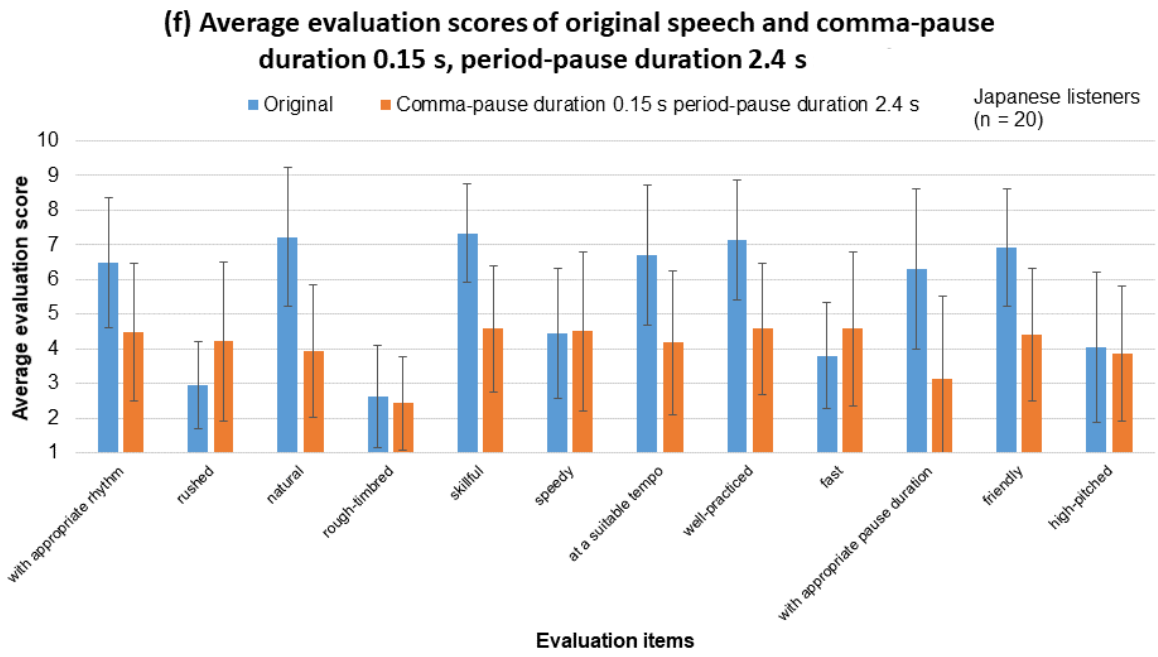


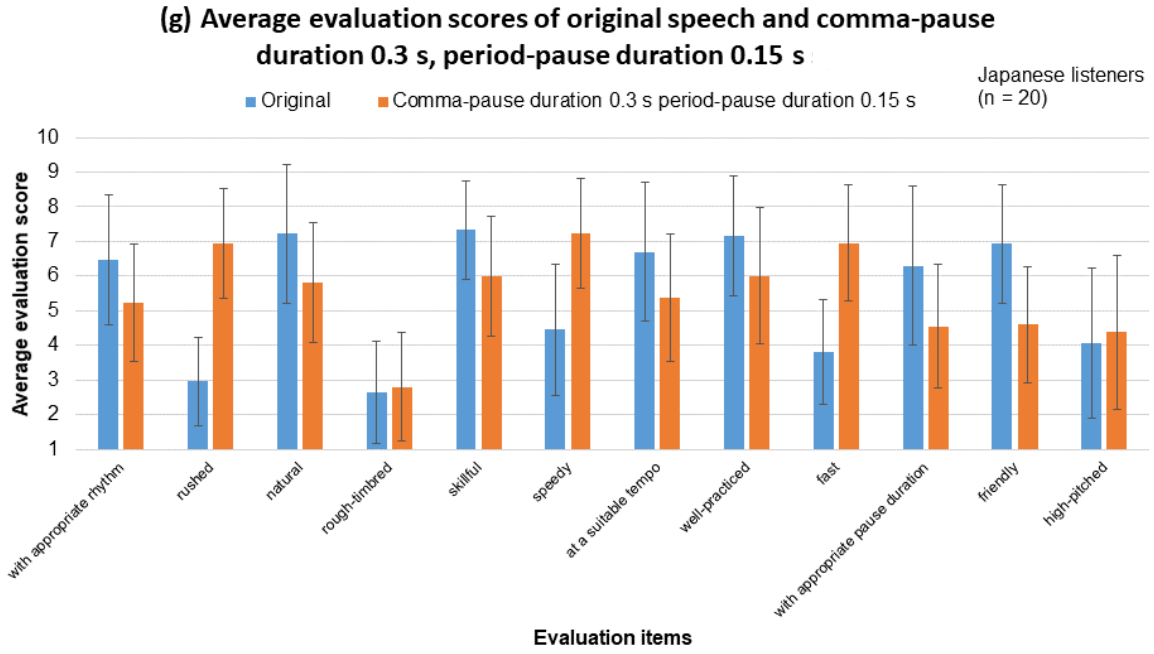


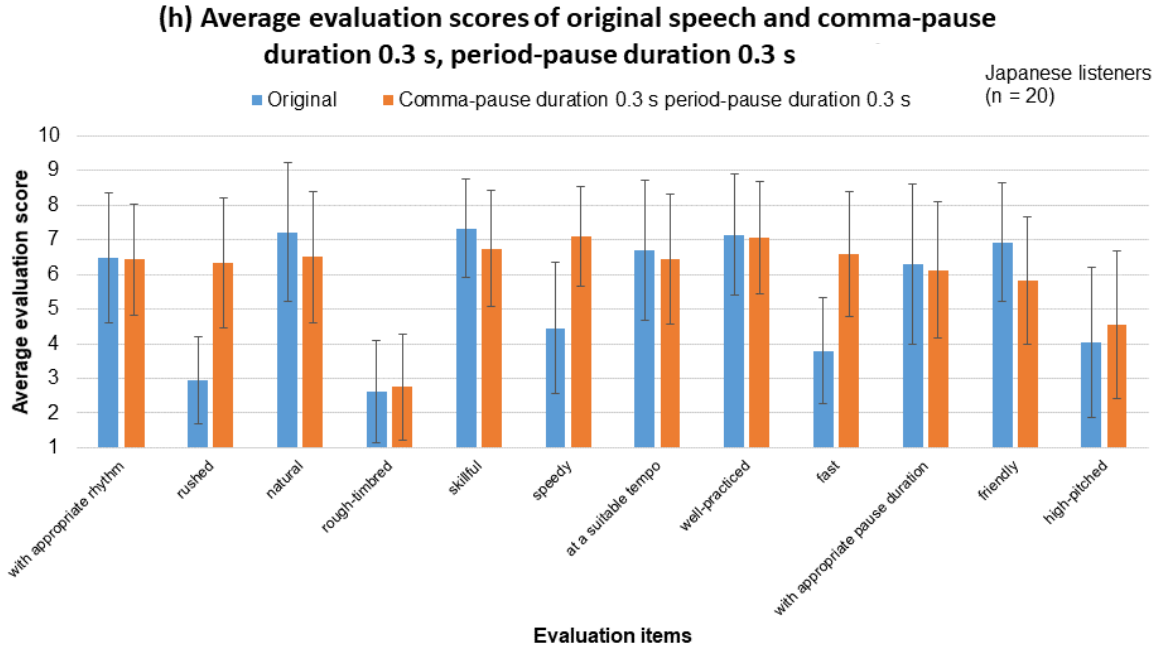

(c) Average evaluation scores of original speech and comma-pause duration 0.15 s, period-pause duration 0.3 s
Japanese listeners (n = 20)
Original
Comma-pause duration 0.15 s period-pause duration 0.3 s
Average evaluation score
10
9
8
7
6
5
4
3
2
1
with appropriate rhythm
rushed
natural
rough-timbred
skillful
speedy
at a suitable tempo
well-practiced
fast
with appropriate pause duration
friendly
high-pitched
Evaluation items

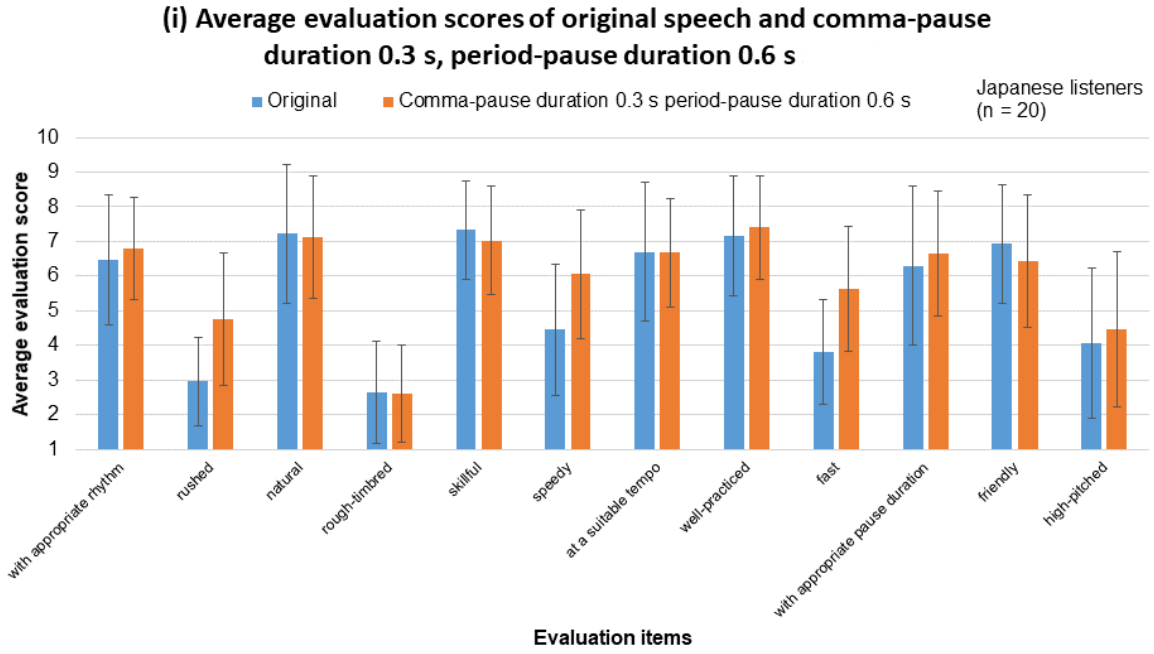

(d) Average evaluation scores of original speech and comma-pause duration 0.15 s, period-pause duration 0.6 s
Japanese listeners (n = 20)
Original
Comma-pause duration 0.15 s period-pause duration 0.6 s
Average evaluation score
10
9
8
7
6
5
4
3
2
1
with appropriate rhythm
rushed
natural
rough-timbred
skillful
speedy
at a suitable tempo
well-practiced
fast
with appropriate pause duration
friendly
high-pitched
Evaluation items

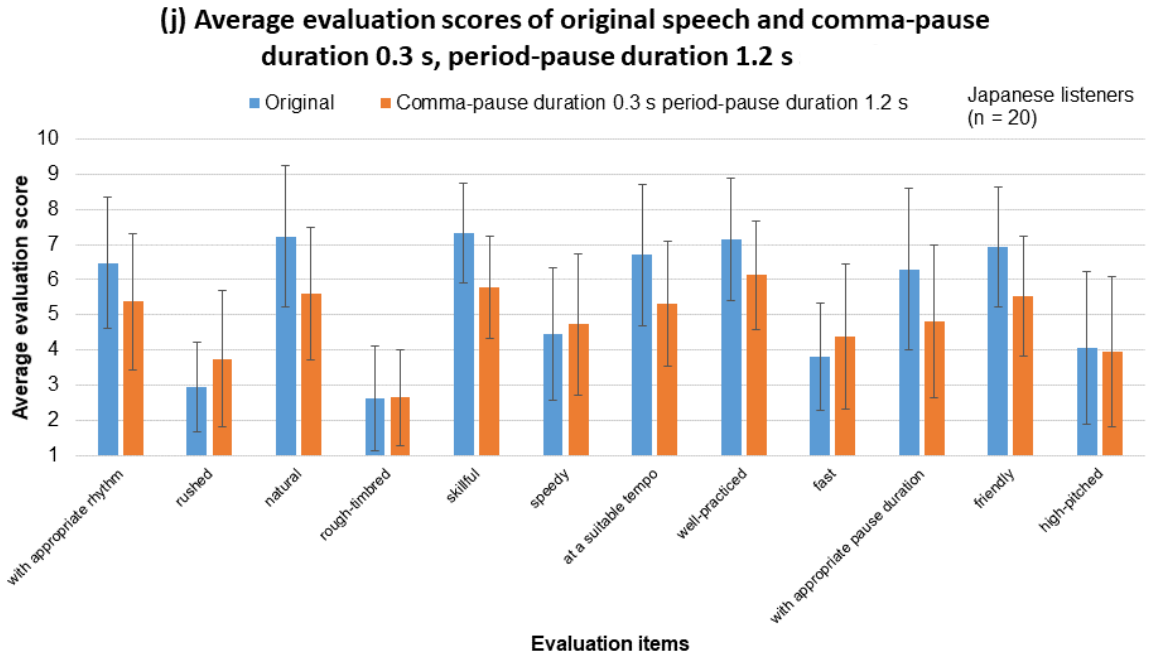

(e) Average evaluation scores of original speech and comma-pause duration 0.15 s, period-pause duration 1.2 s
Japanese listeners (n = 20)
Original
Comma-pause duration 0.15 s period-pause duration 1.2 s
Average evaluation score
10
9
8
7
6
5
4
3
2
1
with appropriate rhythm
rushed
natural
rough-timbred
skillful
speedy
at a suitable tempo
well-practiced
fast
with appropriate pause duration
friendly
high-pitched
Evaluation items


(f) Average evaluation scores of original speech and comma-pause duration 0.15 s, period-pause duration 2.4 s
Japanese listeners (n = 20)
Original
Comma-pause duration 0.15 s period-pause duration 2.4 s
Average evaluation score
10
9
8
7
6
5
4
3
2
1
with appropriate rhythm
rushed
natural
rough-timbred
skillful
speedy
at a suitable tempo
well-practiced
fast
with appropriate pause duration
friendly
high-pitched
Evaluation items

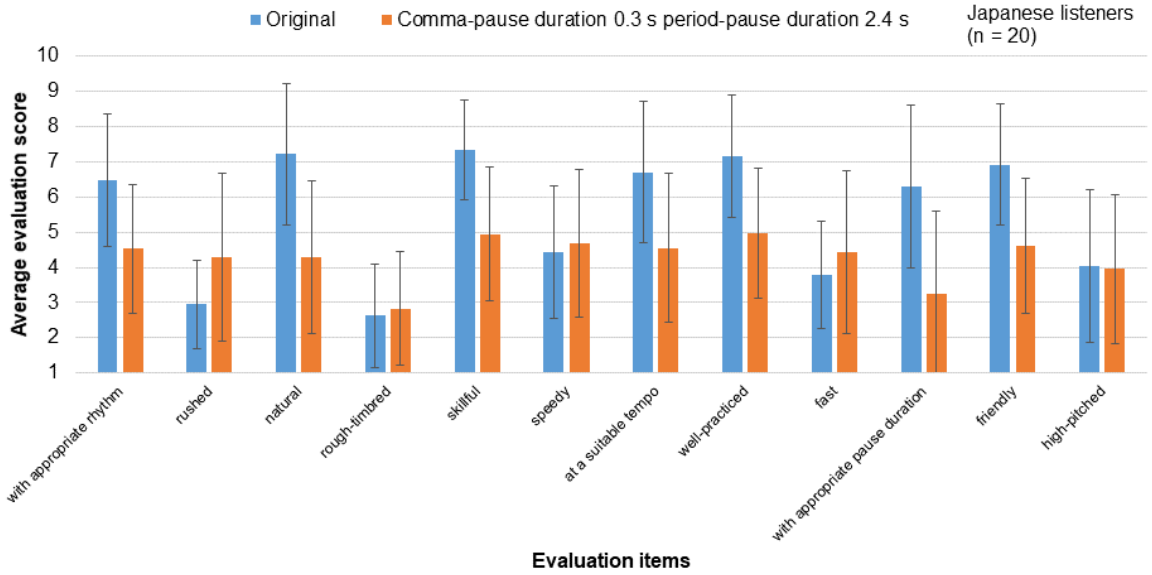

(g) Average evaluation scores of original speech and comma-pause duration 0.3 s, period-pause duration 0.15 s
Japanese listeners (n = 20)
Original
Comma-pause duration 0.3 s period-pause duration 0.15 s
Average evaluation score
10
9
8
7
6
5
4
3
2
1
with appropriate rhythm
rushed
natural
rough-timbred
skillful
speedy
at a suitable tempo
well-practiced
fast
with appropriate pause duration
friendly
high-pitched
Evaluation items


(h) Average evaluation scores of original speech and comma-pause duration 0.3 s, period-pause duration 0.3 s
Japanese listeners (n = 20)
Original
Comma-pause duration 0.3 s period-pause duration 0.3 s
Average evaluation score
10
9
8
7
6
5
4
3
2
1
with appropriate rhythm
rushed
natural
rough-timbred
skillful
speedy
at a suitable tempo
well-practiced
fast
with appropriate pause duration
friendly
high-pitched
Evaluation items

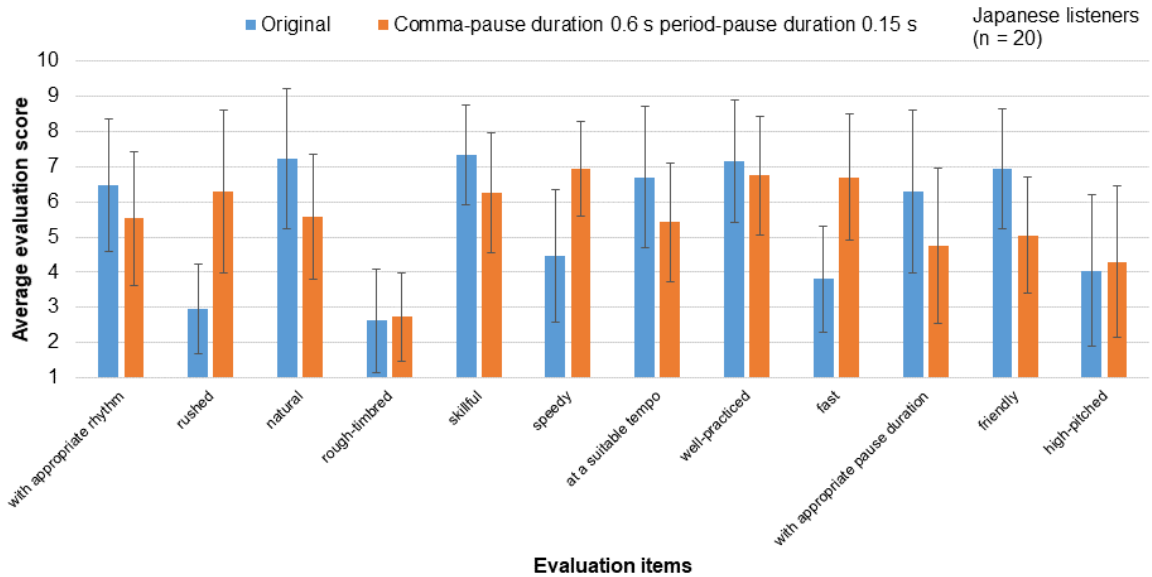

(i) Average evaluation scores of original speech and comma-pause duration 0.3 s, period-pause duration 0.6 s
Original
Comma-pause duration 0.3 s period-pause duration 0.6 s
Japanese listeners (n = 20)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
with appropriate rhythm
rushed
natural
rough-timbred
skillful
speedy
at a suitable tempo
well-practiced
fast
with appropriate pause duration
friendly
high-pitched
Evaluation items

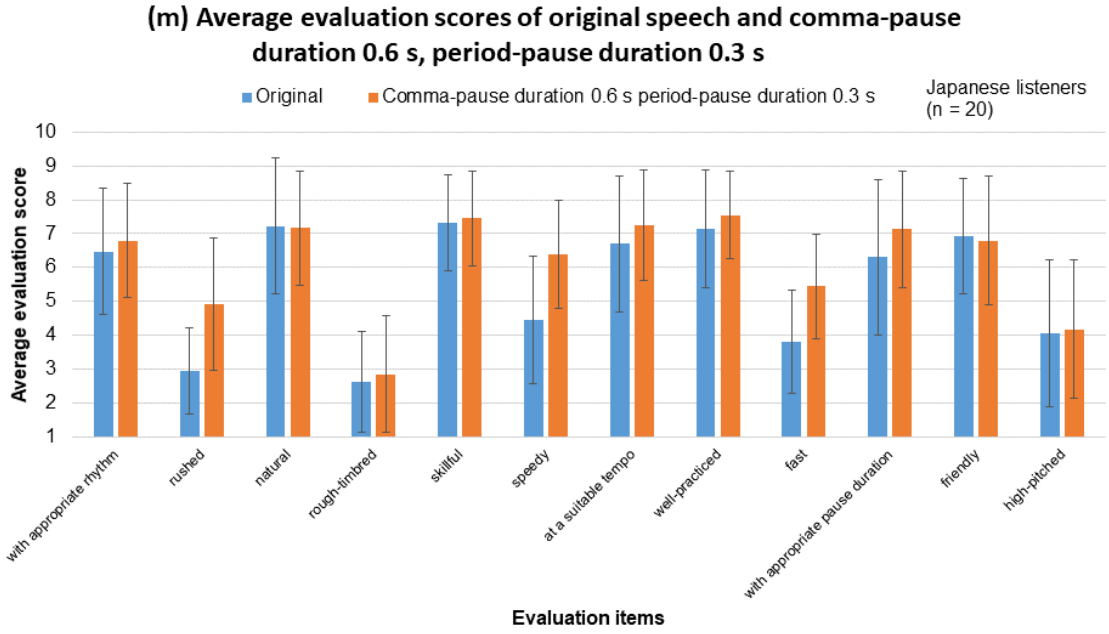

(j) Average evaluation scores of original speech and comma-pause duration 0.3 s, period-pause duration 1.2 s
Original
Comma-pause duration 0.3 s period-pause duration 1.2 s
Japanese listeners (n = 20)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
with appropriate rhythm
rushed
natural
rough-timbred
skillful
speedy
at a suitable tempo
well-practiced
fast
with appropriate pause duration
friendly
high-pitched
Evaluation items

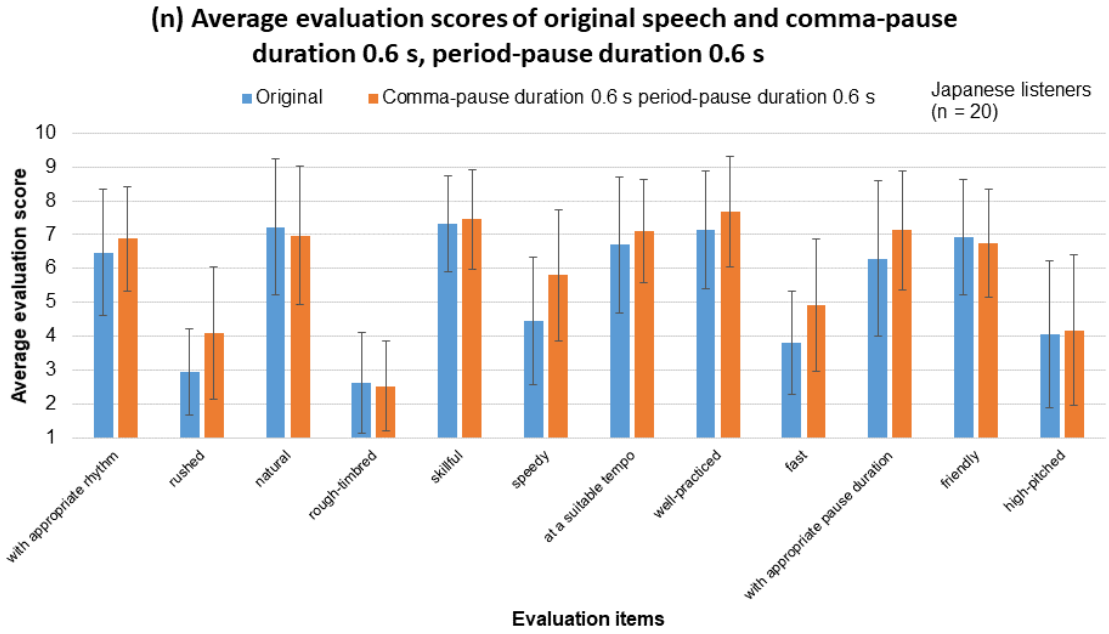
(k) Average evaluation scores of original speech and comma-pause duration 0.3 s, period-pause duration 2.4 s



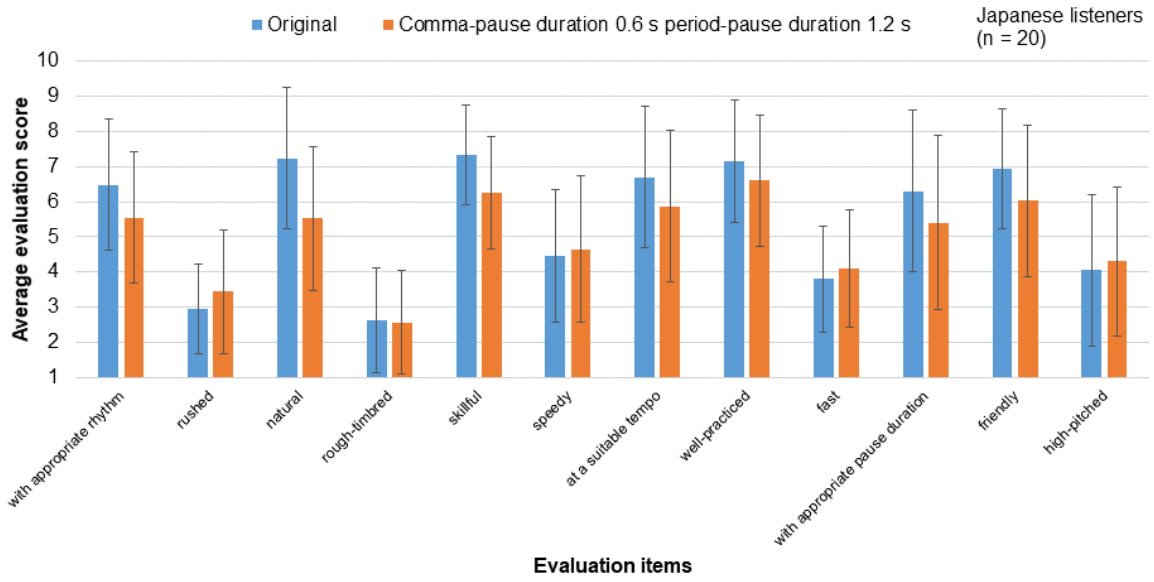
(l) Average evaluation scores of original speech and comma-pause duration 0.6 s, period-pause duration 0.15 s



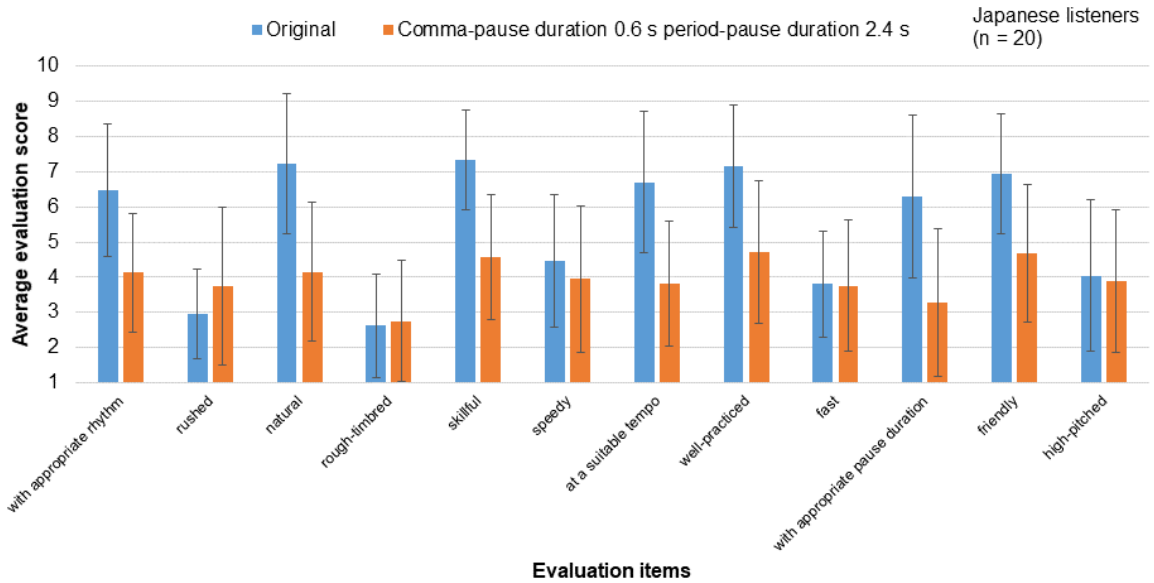

(m) Average evaluation scores of original speech and comma-pause duration 0.6 s, period-pause duration 0.3 s
Original
Comma-pause duration 0.6 s period-pause duration 0.3 s
Japanese listeners (n = 20)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
with appropriate rhythm
rushed
natural
rough-timbred
skillful
speedy
at a suitable tempo
well-practiced
fast
with appropriate pause duration
friendly
high-pitched
Evaluation items


(n) Average evaluation scores of original speech and comma-pause duration 0.6 s, period-pause duration 0.6 s
Original
Comma-pause duration 0.6 s period-pause duration 0.6 s
Japanese listeners (n = 20)
Average evaluation score
10
9
8
7
6
5
4
3
2
1
with appropriate rhythm
rushed
natural
rough-timbred
skillful
speedy
at a suitable tempo
well-practiced
fast
with appropriate pause duration
friendly
high-pitched
Evaluation items

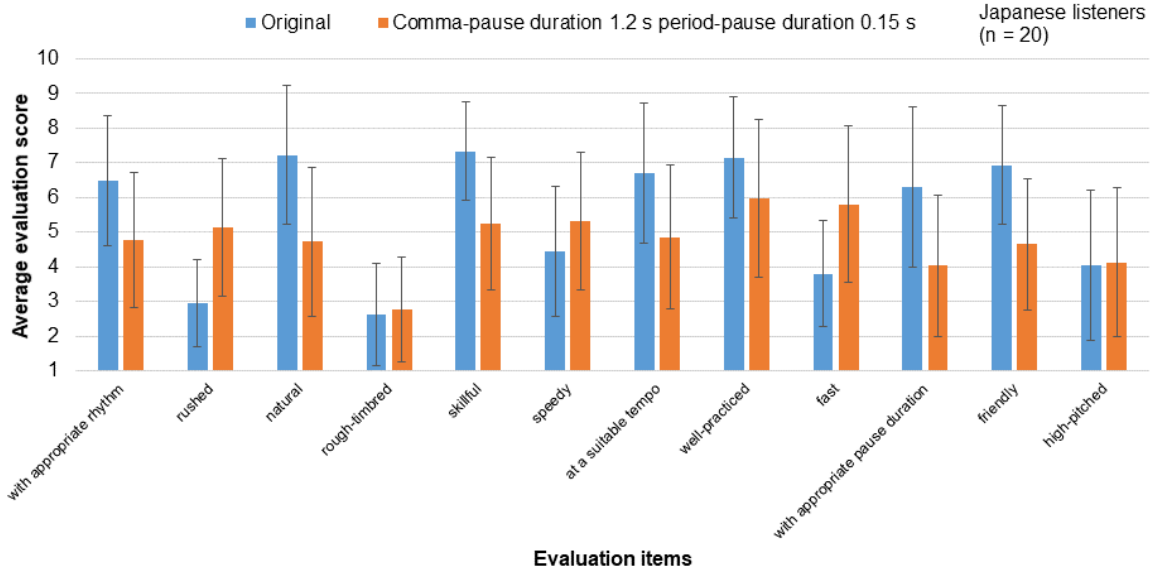
(o) Average evaluation scores of original speech and comma-pause duration 0.6 s, period-pause duration 1.2 s



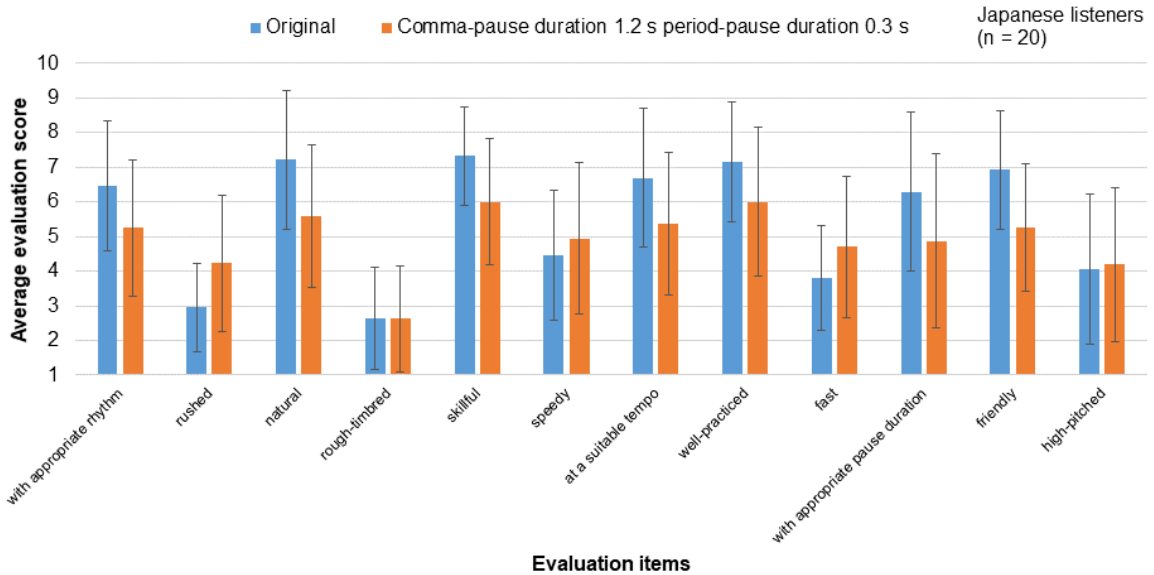
(p) Average evaluation scores of original speech and comma-pause duration 0.6 s, period-pause duration 2.4 s



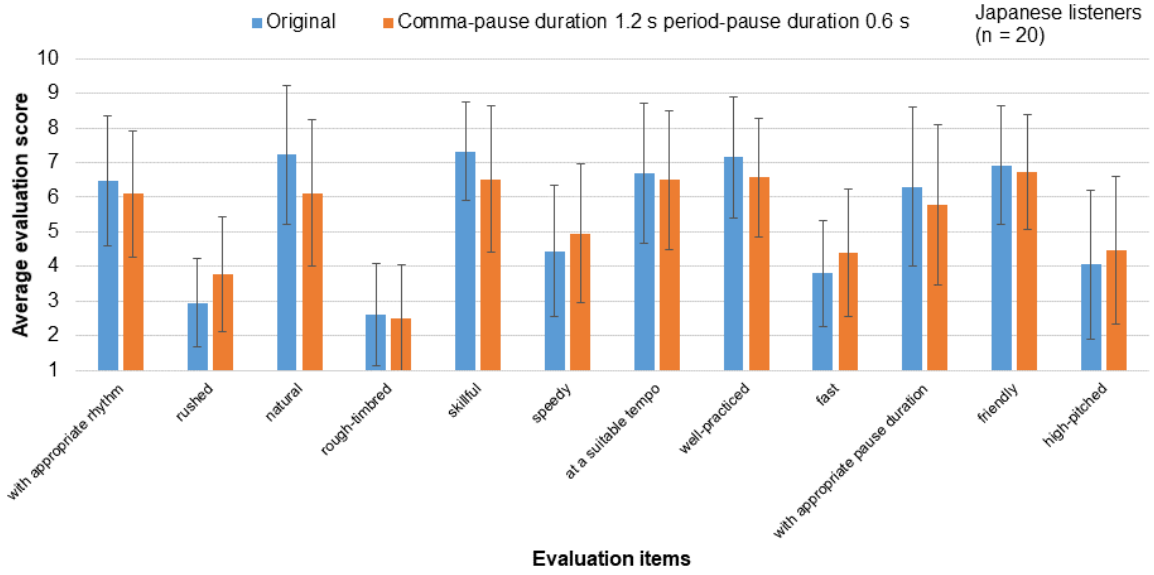
(q) Average evaluation scores of original speech and comma-pause duration 1.2 s, period-pause duration 0.15 s



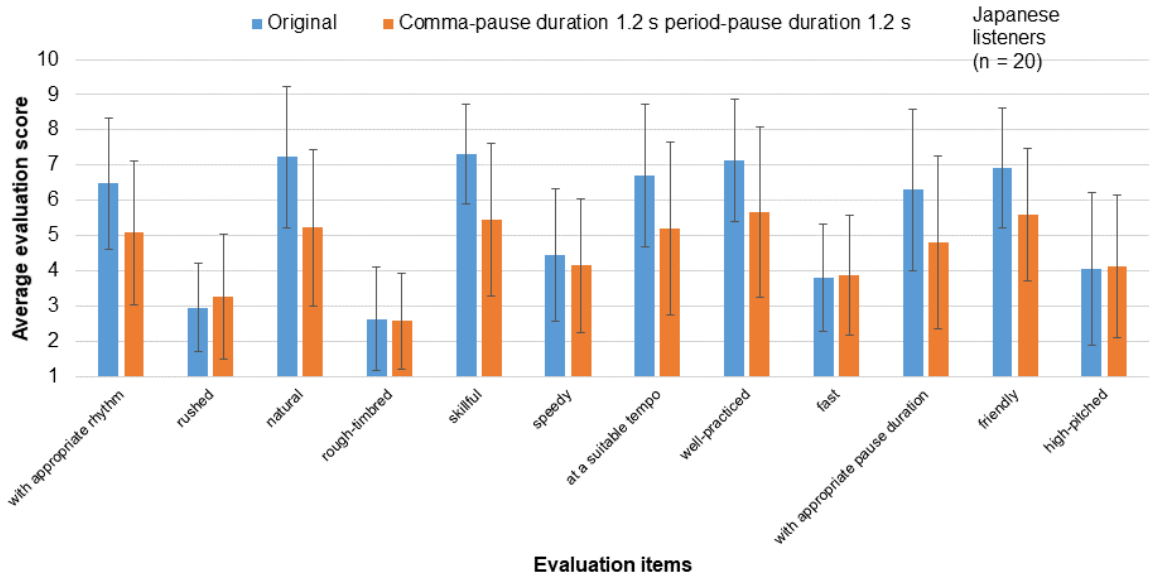
(r) Average evaluation scores of original speech and comma-pause duration 1.2 s, period-pause duration 0.3 s



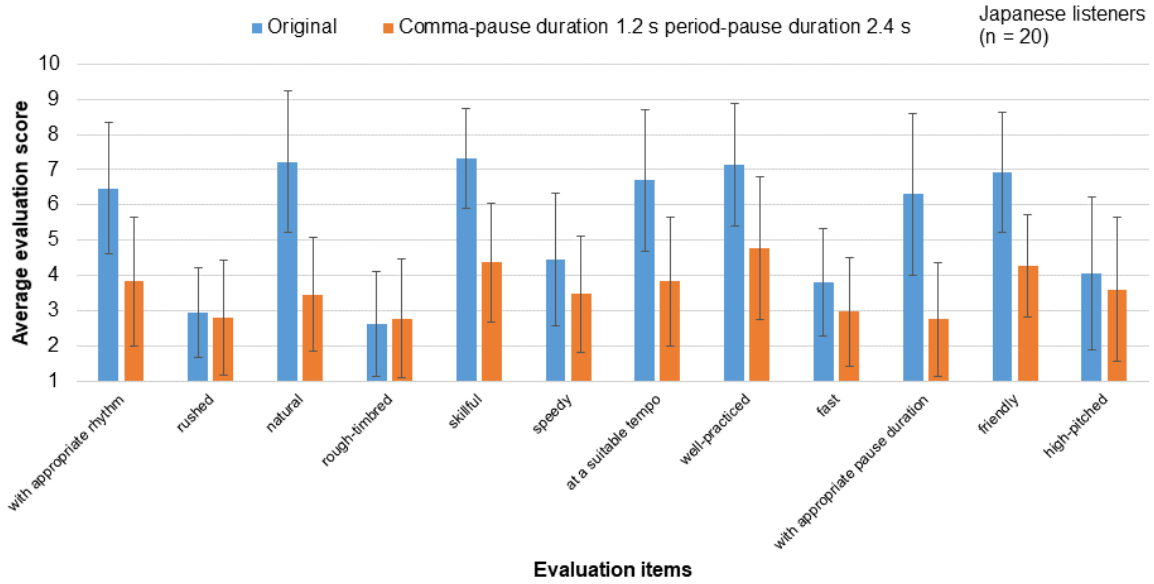
(s) Average evaluation scores of original speech and comma-pause duration 1.2 s, period-pause duration 0.6 s



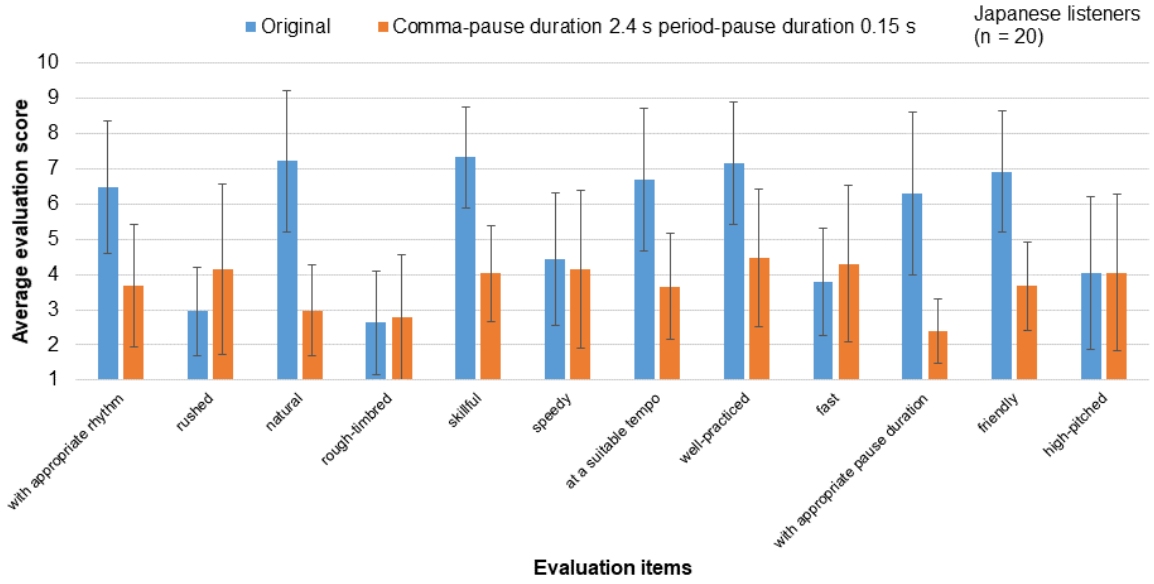
(t) Average evaluation scores of original speech and comma-pause duration 1.2 s, period-pause duration 1.2 s



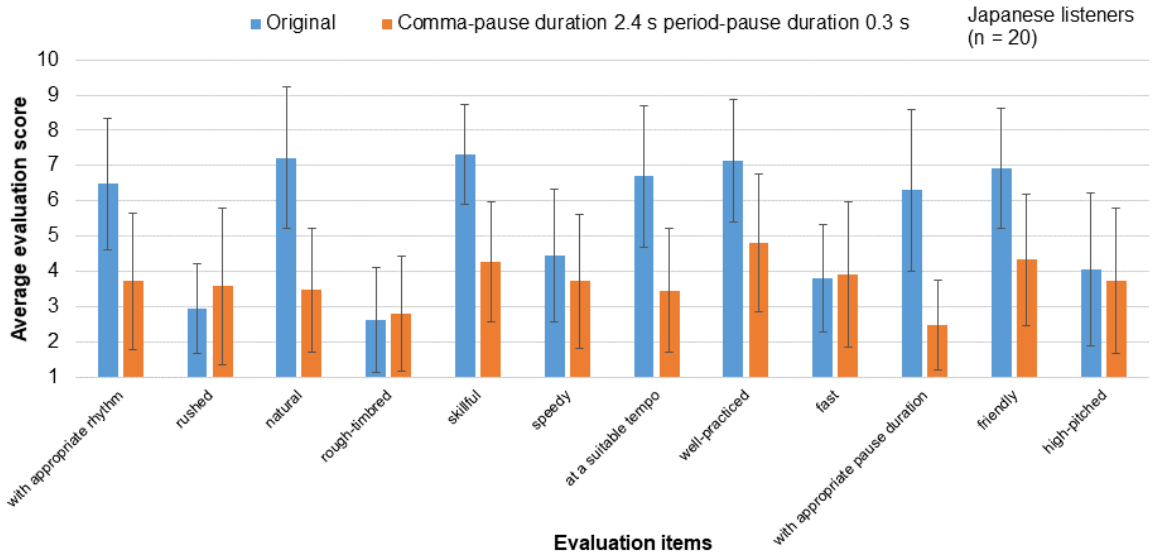
(u) Average evaluation scores of original speech and comma-pause duration 1.2 s, period-pause duration 2.4 s



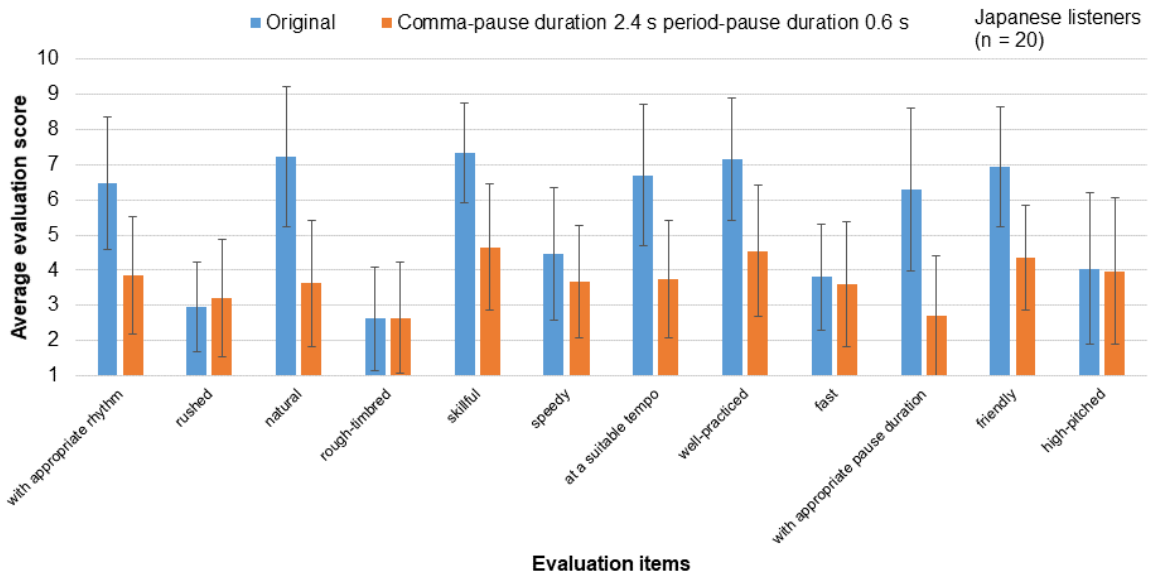
(v) Average evaluation scores of original speech and comma-pause duration 2.4 s, period-pause duration 0.15 s



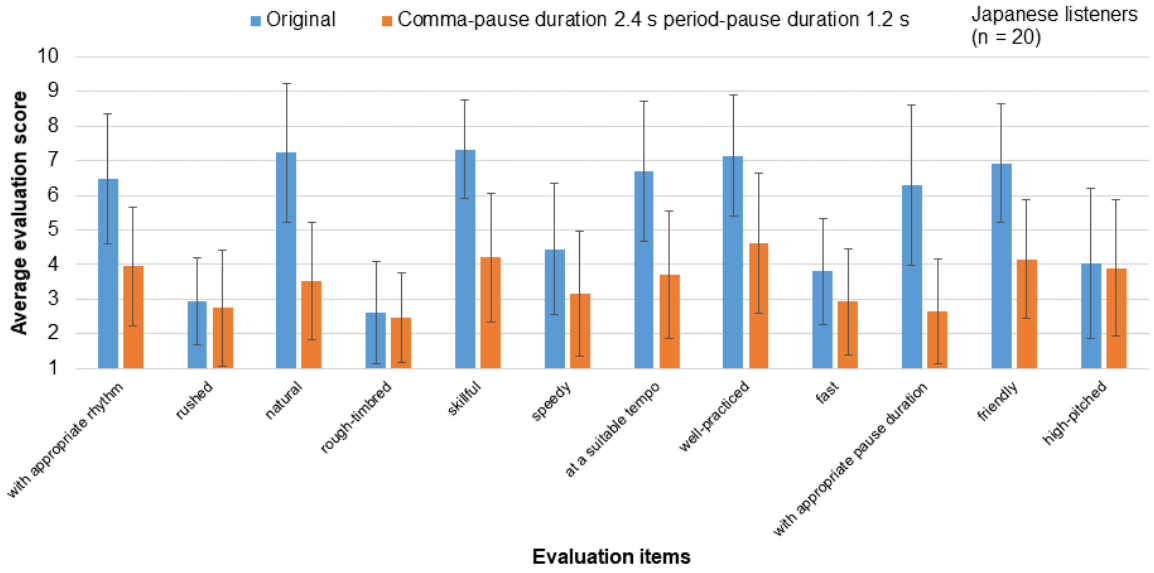
(w) Average evaluation scores of original speech and comma-pause duration 2.4 s, period-pause duration 0.3 s



(x) Average evaluation scores of original speech and comma-pause duration 2.4 s, period-pause duration 0.6 s



(y) Average evaluation scores of original speech and comma-pause duration 2.4 s, period-pause duration 1.2 s



(z) Average evaluation scores of original speech and comma-pause duration 2.4 s, period-pause duration 2.4 s