



Perception and Reinterpretation of English Song Lyrics by Native Speakers of Japanese: A Case Study of Samples From the TV-Show *Soramimi-Hour*

Johannes Scherling^{1*}, Lisa Kornder¹ and Niamh Kelly²

¹ Department of English Studies, University of Graz, Graz, Austria, ² Department of Languages and Linguistics, University of Texas at El Paso, El Paso, TX, United States

OPEN ACCESS

Edited by:

Alexander Onysko,
University of Klagenfurt, Austria

Reviewed by:

Philip Seargeant,
The Open University, United Kingdom
Takeshi Nozawa,
Ritsumeikan University, Japan

*Correspondence:

Johannes Scherling
johannes.scherling@uni-graz.at

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Communication

Received: 20 September 2021

Accepted: 02 March 2022

Published: 15 April 2022

Citation:

Scherling J, Kornder L and Kelly N
(2022) Perception and
Reinterpretation of English Song Lyrics
by Native Speakers of Japanese: A
Case Study of Samples From the
TV-Show *Soramimi-Hour*.
Front. Commun. 7:780279.
doi: 10.3389/fcomm.2022.780279

Cross-linguistic mondegreens occur when foreign song lyrics are misperceived and reinterpreted in the listener's native language. In Japan, such humorous reinterpretations of non-native song lyrics are known as *soramimi* (空耳, "mishearing"). Word plays of this kind do not only have an entertaining character for listeners, but they also offer a valuable source to identify and describe potential phonological processes which can be observed in native Japanese listeners' adaptations of English song lyrics into Japanese. We hypothesized that the reinterpretation of English song lyrics by listeners is a result of the perception of non-native auditory input through the first language's phonological and morphological system. That is, misperceptions do not occur arbitrarily, but are governed by the phonological and morpho-phonological rules of the listener's first language system. To test this hypothesis, we examined a corpus containing 60 English-Japanese mondegreens taken from the Japanese TV-show *Soramimi Awā* (*Soramimi Hour*). Results confirmed our hypothesis: The Japanese adaptations were observed to follow different phonological processes which aimed to subject the non-native auditory input to the phonological rules of Japanese.

Keywords: Japanese, English, cross-linguistic mondegreens, phonological processes, Japanese phonotactics, top-down processing, language contact

INTRODUCTION

In the present paper, we aimed to find out which phonological processes underlie native Japanese listeners' misperceptions and reinterpretations of English song lyrics, known as *soramimi* in Japanese (空耳, "mishearing"). This investigation is based on the hypothesis that native Japanese listeners reinterpret English auditory input through the phonological filter of their native language. In this context, we argue that—similar to the adaptation of English loanwords into Japanese—the misperception of foreign language lyrics is governed by the phonological rules of listeners' first language system. A corpus of 60 English-Japanese *soramimi* was analyzed and, based on a comparison of the English original and the Japanese misperceived lyrics, categorized according to one of four (morpho-)phonological processes, that is, sound substitution, vowel insertion, segment deletion, and boundary transgression. After briefly outlining the history of language

contact between Japanese and English, aspects that relate to Japanese phonotactics will be discussed. In Section Integration Processes for Foreign Language Material in Japanese, integration processes which are relevant when it comes to adapting foreign language material into Japanese are described. This is followed by an outline of English phonotactics (Section English Phonotactics) and briefly comparing the English and Japanese vowel and consonant inventories (Section The English and Japanese Sound Inventories). Section Mondegreens, *Soramimi*, and Speech Perception explains what mondegreens and *soramimi* are and how they are related to speech perception. In the subsequent section, we outline our methodology and provide an analysis and discussion of our findings.

A SHORT HISTORY OF LANGUAGE CONTACT BETWEEN JAPANESE AND ENGLISH

Though language contact in Japan has been a comparatively rare occurrence due to its geography as an island nation, when it did happen, it affected the Japanese language considerably. The best example for this is the intensive contact between the Japanese and the Chinese language that took place from the first century AD (Hoffer, 2002, p. 29; Loveday, 1996, p. 264-65). This profound cultural exchange not only led to the introduction of Buddhism by the sixth century—mostly *via* Sino-Korean influence (Loveday, 1996, p. 39-40)—but also resulted in the wholesale adoption of the Chinese writing system and lexicon [~50% of the contemporary Japanese lexicon are of Chinese origin (Morrow, 1987; Stanlaw, 2004; Scherling, 2012)], which over time the Japanese “were able to develop and rework [...] according to their own needs” (Loveday, 1996). Out of these Chinese characters, Japanese then developed its two syllabaries, namely Hiragana and Katakana (Matsuda, 1985; Stanlaw, 2004), in order to enable people to read texts without having to be able to read the complex Chinese characters only known to the learned at that time. These syllabaries represented all possible sound combinations known to Japanese. Applying the phonotactic constraints determined by the syllabaries to foreign words then became the blueprint for assimilative processes that—alongside morphological processes such as clipping and blending, based on the Chinese model, where loanwords are treated as uninflected and bound bases (Loveday, 1996, p. 138) – have shaped loanword integration¹ in Japan to this day (Stanlaw, 2004).

Contact to European languages came rather late. Only when Europe’s empires became seaborne did a chance for language contact arise. The first to reach the shores of Japan—in the sixteenth and seventeenth century—were the Portuguese, followed by the Spanish and the Dutch, all of which left behind a number of loanwords still in use today, such as “pan” (パン, Portuguese for “bread”), “kasutera” (カステラ, Spanish for

“sponge cake”) or “garasu” (ガラス, Dutch for “glass”). Japan then delved into a self-imposed period of isolation, a result of arguably well-founded fears of meddling or even colonization (Stanlaw, 2004, p. 46-47). It took Japan some 250 years before opening up to the world again—under the threat of force. In 1853, US-American Commodore Matthew Perry sailed into Edo Bay (what is nowadays Yokohama Bay) with heavily armed gunships to successfully force Japan to open its ports to international trade again (LaFeber, 1997, p. 13). It was through this traumatic experience that the Japanese government realized its lag in technological advancement to the world and felt the need to “modernize” and “westernize” by adopting and assimilating all things Western (Loveday, 1996, p. 62; Stanlaw, 2004, p. 52-56). Increasingly, then, by the early twentieth century, the means of “modernization” became the English language, with words like “takushi-” (タクシー, Eng. “taxi”), “rajio”, (ラジオ, Eng. “radio”) or “jiruba” (ジルバ, Eng. “jitterbug”) entering the language in quick succession (Stanlaw, 2004). After a short, but all the more intense intermezzo of hostilities during the 1930s and 1940s which culminated in the Pacific War, during which English loanwords were partially replaced by words of Chinese and Japanese origin (Ōishi, 1992), and after Japan’s defeat, epitomized through the devastation of two atomic bombs, the English language celebrated a phenomenal comeback; this time as the perceived key to attaining the much admired living standards of the “American Way of Life” (Loveday, 1996; Dower, 1999). In the aftermath of the war, the number of English-based loanwords started to increase steeply, reaching proportions of more than 10% of the entire Japanese lexicon (Scherling, 2012), a number that might be even higher if the many hybrid words (such as “shirobai” (白バイ), meaning “motorcycle policeman”, a combination of “shiroi”—白い, “white”—and “bike”) that contain English language material are taken into account.

One consequence of this modernization—and a gradual but incremental effect leading back as far as the Meiji Restoration of 1868 (Pintér, 2015, p. 175)—has been innovations in the Japanese syllabary system. These became necessary in order to take into account as of yet unknown sounds and sound sequences that were needed to assimilate foreign words into the Japanese linguistic matrix, such as “fi” (ファイ), as in “fire” (ファイレ, Eng. “filet”) which had previously been rendered as “hire” (ヒレ), or “va” (ヴァ), as in “vaorin” (ヴァイオリン, Eng. “violin”) which used to be “baiorin” (バイオリン) (Stanlaw, 2002). These innovations have undoubtedly made it easier to adapt English words into Japanese phonology. They have, however, changed little regarding the very basic phonological differences between the two languages, which instantly necessitate adaptations in the sound structure of English-based loanwords on their arrival in Japanese and which will be the focus of the following section.

JAPANESE PHONOTACTICS

The Japanese language is—with the exception of some more differentiated views (Warner and Arai, 2001)—widely regarded

¹From a phonetic/phonological perspective, loanword adaption can be described as “a process that applies during speech perception and that maps non-native sound structures onto the phonetically closest native ones” (Peperkamp, 2005, p. 2).

a mora-timed language “where each mora is supposed to take an equal duration of time” (Kubozono, 2002, p. 33). A mora is a phonological unit describing syllable weight (Hogg, 1992) and is also called a phonemic syllable (e.g., Pike, 1947; Kubozono, 2002). It differs from the syllable in that, while every syllable consists of (at least) one mora, not every mora can constitute a syllable (Kubozono, 2015a, p. 63). For instance, while “Nagasaki” consists both of 4 syllables and 4 morae, a word like “Tōkyō” has only 2 syllables but 4 morae, due to the two long vowels (essentially To-o-kyo-o). For the present investigation, understanding the importance of morae is vital as it “plays a crucial role in speech perception” (Kubozono, 2002, p. 39) and because Japanese speakers have been shown to “respond to mora-sized units more readily than syllable-sized units” (Kubozono, 2002 citing Hayashi and Kakehi, 1990).

Owing to its syllabic structure (reflected in its writing system which requires consonants to be represented along with a vowel and not individually, while vowels can be represented individually), the phonotactic constraints that apply to Japanese words or words that are adopted into Japanese are rather rigid. Japanese permits only open syllables, with the sole exception of the nasal /N/. This means that the only permissible sound sequences are (C)V, (C)VV, and (C)VN (Dupoux et al., 1999, p. 2), with the caveat that not all consonants can combine with all vowels (Pintér, 2015, p. 174). For example, the consonant /t/ can combine with /a/, /e/ and /o/, but not with /i/ and /u/, whereas /s/ cannot combine with /i/, and /w/ can only combine with /a/. Despite the innovative changes to the syllabary in recent decades (see above), this constrains and influences the integration of non-Japanese words into the writing system, in particular since, as Stanlaw suggests, there is a recent trend to use the more conservative spelling, and the new syllabary characters are not all that widespread (Stanlaw, 2004, p. 95 citing Inoue, 1996, p. 196-97).

Therefore, when novel words enter the Japanese language from another language, the first impact they experience is on their phonology, as many sound combinations from languages such as German or English are not permissible in Japanese and can therefore not be represented without some changes. Since Japanese, as outlined above, disallows consonant clusters or closed syllables, such consonant clusters need to be broken up or complemented with vowels so that they can fit Japanese phonological/phonotactic structure. Another theory maintained by Dupoux et al. is that, in fact, Japanese phonotactic constraints lead Japanese speakers to “perceive epenthetic [u] vowels within consonant clusters” (Dupoux et al., 1999, p. 11) to adapt foreign language input to their language’s phonotactics. The exact phonological processes which are relevant in this context will be described in more detail below.

Another important aspect of Japanese phonology pertains to the production of vowels in certain environments, a phenomenon that is called “devoicing”. Devoicing means that vowels that are typically pronounced with the vocal folds vibrating under certain conditions lose that voicing, one of those conditions being speaking rate (Fujimoto, 2015, p. 215). According to Fujimoto, devoicing affects the high vowels /i/

and /u/ in the morae /pi/, /pu/, /ki/, /ku/, /ci/,² /cu/, /tei/, /tɕu/, /hi/, /su/, /tsu/, and /hu/, in particular when they are followed by a voiceless consonant (ibid). Hence, a word such as “shita” (下, “below”) would tend to be pronounced like [çita], or “shuto” (首都, “capital”) like [çuto], to the extent that the vowel is barely perceivable anymore. Devoicing also applies in word-final environments, such as “desu” (です, “to be”), which is pronounced like [desɯ]. Acoustic examinations, however, confirm that, despite the devoicing and possible subsequent reduction in distinction between /i/ and /u/, native Japanese speakers are both able to differentiate between and identify different vowels with identical onset consonants (e.g., /ci/ and /cɯ/) (Fujimoto, 2015, p. 194). In this respect, what is of interest in the context of the present study is whether native Japanese listeners will show a tendency to “hear” devoiced vowels in consonant clusters from English words in their “misheard” song lyrics as well, seeing that speaking rate in songs tends to be faster the faster the melody is (Jungers et al., 2002) and given that “faster speech facilitates devoicing in non-general conditions and atypical consonantal conditions” (Fujimoto, 2015, p. 206).

INTEGRATION PROCESSES FOR FOREIGN LANGUAGE MATERIAL IN JAPANESE

As discussed above, the particular phonotactic rules and constraints of the Japanese language necessitate fundamental changes to the phonological structure of loanwords. This is a result of its syllabary alphabet which cannot, for example, represent individual consonants or consonant clusters. According to Loveday (1996, p. 114), on arrival in Japanese, “the syllabic structure of English is radically altered and new sets of morpho-phonemic and phonotactic patterns are introduced that are not always regular or predictable”. In the following, the most frequent phonological processes that such loanwords undergo when adopted will be discussed with a view to mishearings of English song lyrics, which we argue are subject to similar strategies and processes.

The two most frequent processes affecting loanword integration are sound substitutions and vowel epenthesis, or insertion (Loveday, 1996; Nian and Jubilado, 2011; Scherling, 2013). Sound deletion also plays a certain role, in particular for unstressed syllables or in environments where a consonant coda is not perceived or not realized (Shoji and Shoji, 2014). Vowel epenthesis, as argued above, is a natural consequence of Japanese orthography, phonotactics and language perception. When disallowed consonant clusters or consonant codas are encountered, these are broken up by inserting vowels, mostly /i/, /o/ and /u/, to ensure open syllables. According to Stanlaw (2004, p. 74), the choice of the epenthetic vowel is a function

²The source referred to here, but also others we cite in this paper, use the IPA symbols /ʃ/, /ʒ/, /tʃ/ and /dʒ/ to refer to Japanese alveolo-palatal fricatives and affricates. As Kubozono (2015a, p. 6) states, this is debated but not uncommon, yet for the sake of consistency with our own transcriptions of the respective consonants (/ç/, /z/, /tç/, /dʒ/) we have decided in favor of replacing the IPA symbols used in the original transcriptions with the ones we employ in ours.

of the phonological environment: /u/, called the “context-free default epenthetic vowel” by Shoji and Shoji (2014, p. 3), is inserted after most consonants. While /o/ usually follows after the alveolar consonants /t/ and /d/, /i/ is added after affricates /tʃ/ and /dʒ/—the two being “context-dependent epenthetic vowels” (Shoji and Shoji, 2014, p. 3). Therefore, an English word such as “infrastructure” would be rendered as “infurasutorakucha-” (インフラストラクチャー) in Japanese, a word like “bridge” would become “buriiji” (ブリッジ, where the duplicated “j” is a geminated voiced affricate), and “drugstore” would become “doragusutoa” (ドラグストア). Clearly, such inserted vowels make the loanword longer in terms of syllables and also incrementally remove it phonologically from the source word, so that they would become difficult to identify as English words for native English speakers. However, as Dupoux et al. (1999) have argued, for Japanese speakers, such vowels appear to be actually perceived, which may help explain some of the mondegreens to be discussed later.

The second process, substitution, applies to vowels and consonants that are non-phonemic in Japanese and therefore need to be replaced with phonologically similar, approximate sounds available in the Japanese sound inventory. There is a considerable number of sounds that are unknown in Japanese, such as the interdental fricatives /θ/ and /ð/, the lateral approximants /r/ and /l/, and the labiodental fricative /v/. There are also illicit syllables like /tɪ/ or /dɪ/, as well as non-existent vowels, such as /ɛ/ or /æ/ (Matsuda, 1986, p. 49; Stanlaw, 2004, p. 74). Foreign words that contain such sounds are subject to sound substitution. The choice of the sound that will substitute for the non-phonemic one is variable and considerably depends on the phonological environment. As previously mentioned, not all consonants combine with all vowels in Japanese. Hence, the interdental /θ/ can be substituted by sounds as different as /s/ (third > sādo), /tʃ/ (thick > shikku) or /ts/ (thulium > tsuriumu) (Stanlaw, 2004, p. 74). This is because the combination /sɪ/ is disallowed in Japanese, and must be realized as /tʃɪ/. For “thulium” (see above), the reason is possibly that the word was imported from German rather than English and hence the /tu/ sound sequence had to be replaced by /tsu/ as it is not a permitted sound combination in Japanese. For vowels, the patterns are more regular, but somewhat dependent on whether the perception of the foreign word happens visually or aurally³. An example of that is “Christmas” (/krɪsməs/), in which case the word was likely first encountered visually, since the hardly perceptible mid central vowel /ə/ is rendered as /a/ in Japanese (/kurusumasu/). Nian and Jubilado (2011, p. 101) and Tsuchida

(1995, p. 147-48), for instance, outline the following broad regularities when it comes to the substitution of vowels:

Eng. /ɜ/ > Jpn. /a/: e.g., “bird” /bɜ:d/ > bādo /ba:do/
 Eng. /æ/ > Jpn. /a/: e.g., “taxi” /tæksi/ > takushi- /takuʃi:/
 Eng. /ɛ/ > Jpn. /e/: e.g., “pet” /pɛt/ > petto /petto/
 Eng. /ɔ/ > Jpn. /o/: e.g., “boarding” /bɔ:rdɪŋ/ > bodingu /bo:dingu/

Similarly, English diphthongs—considering that most scholars agree that diphthongs do not exist in Japanese⁴—are also replaced, for example by long monophthongs, such as “cake” (/keɪk/) becoming “ke-ki” (ケーキ, /ke:ki/) or “shake” (/ʃeɪk/) realized as “she-ku” (シェーク, /ʃe:ku/) (Shoji and Shoji, 2014, p. 6). As with consonants, however, there is also a certain flexibility with how vowels are represented in Japanese, largely dependent on the English variant that the word in question was provided by. Hence, in words such as “ballet” (AE /bæleɪ/) and “volleyball” (AE /vɒlibɔ:l/), the different vowels in English will be represented by the same vowel (/a/) in Japanese: both are rendered as “bare-” (バレー, /bare:/), with the two different consonants /b/ and /v/ being collapsed into one (/b/). Clearly, substitution and vowel epenthesis are closely interlinked, as the choice of the substituting sound may determine the nature of the epenthetic vowel based on the phonotactic constraints of the language. These underlying rules, in turn, may then affect the perception of the sound structure of non-Japanese words.

The last phonological process to be discussed here is sound deletion. As Shoji and Shoji (2014, p. 3) state, Japanese loanwords “prefer epenthesis to deletion,” yet deletion is known to occur, for example, with inflectional suffixes in English, such as “smoked salmon”, which becomes Japanese “sumo-ku sa-mon” (スモークサーモン), or similarly “corned beef”, which is Japanese “kon bi-fu” (コンビーフ) (Shoji and Shoji, 2014, p. 9). The same applies to English “skiing” which is realized in Japanese as “suki-” (スキー), “salaried man” which becomes “sarariman” (サラリーマン), or the phrase “three strikes” which is Japanese “suri- sutoraiku” (スリーストライク) (Stanlaw, 2004, p. 75). In other cases, final consonants may become the target of deletion, such as in the Japanese renderings of “alright” as “o-rai” (オーライ) or “don’t mind” as “donmai” (ドンマイ), as well as “handkerchief” as “hankachi” (ハンカチ). Shoji and Shoji (2014, p. 9) argue that such deletions may be due to a lack of perception of the consonant coda of the closed syllable. Here, an argument can be made that these phonological processes, being related specifically to the mental perception of foreign words through the filter of Japanese phonotactics, apply not only to loanword integration, but more broadly to foreign language perception in general, and may thus also apply to the perception of song lyrics in languages other than Japanese. They further suggest that the choice between epenthesis and deletion often depends on whether the language input occurred *via* the aural or the visual channel, an argument that Stanlaw also proposes when he distinguishes between borrowing by eye

³Previous research suggests that orthography plays a crucial role in loanword adaptations from one language to another (e.g., Kaneko, 2006; Venedlin and Peperkamp, 2006). Kaneko (2006), for instance, investigated the role of orthography in the selection of Japanese vowels in English loanwords and found that presenting phonological forms together with their spelling indeed had an impact on which vowels native Japanese listeners were likely to select. For example, in an oral condition (i.e., presenting phonological forms without spelling), listeners adapted the English low back vowel [ɑ] as the phonetically closest Japanese targets [a] or [a:]. However, if presented in a mixed condition (i.e., phonological forms with spelling), listeners were more likely to adapt the English target vowel as Japanese [o] due to orthographic influences.

⁴Some scholars argue that there are vowel sequences in modern Japanese which are acoustically similar to diphthongs; this particularly applies to the vowel sequences /ai/, /oi/, and /ui/ (see e.g., Kubozono, 2005).

and borrowing by ear (Stanlaw, 2004, p. 91-2). For example, the English word “pudding” exists in Japanese both as a borrowing by eye (“pudingū”, プディング) as well as a borrowing by ear (“purin”, プリン), the second of which shows a case of consonant deletion. Similarly, this applies to loanwords such as “pokke” (ポッケ, Eng. “pocket”), where the /t/ was deleted, or “hankachi” (ハンカチ, Eng. “handkerchief”), where word-final /f/ was omitted (Shoji and Shoji, 2014, p. 9). Hence, aural perception—as is the case in *soramimi*—appears to make deletion more frequent than it is when words are encountered *via* the written mode.

We wish to raise one last issue here, which we are proposing may happen in foreign language perception. It connects the process of deletion of final consonants in the perception of foreign words with a morphological issue that may arise in the perception of longer and uninterrupted stretches of spoken language, such as in songs, and which is referred to as boundary transgressions. Final consonant deletion is possible in isolated, individual words because the consonant coda is not perceived when the word is pronounced out of context with other words. In connected speech, however, when the word in question is followed by another word, the final consonant may well be perceived as being the onset of the following word, in particular when this onset is realized by a vowel. This means that if an English word ends in a consonant coda, which may well be deleted if perceived in isolation, this consonant might—instead of being omitted—be reanalyzed as the onset of the following word, thus potentially leading to a mishearing. An example for this taken from our corpus would be the perception of the English phrase “Nothin’ can hold us and” as “Naze ka hon dashita” (meaning “For some reason, I published a book”), where the final consonant of “hold” was reanalyzed as the initial consonant of the following word “dashita”.

ENGLISH PHONOTACTICS

The phonotactic patterns of English allow complex syllable structures involving consonant clusters. Consonant clusters are licit in onsets as well as in codas, that is, English syllables can be open or closed. Onset clusters generally have a rise in sonority toward the syllable nucleus (although clusters such as /st/ are exceptions to this generalization), while coda clusters have a fall in sonority from the nucleus (e.g., Zec, 2007). As such, English allows more complex syllables than Japanese, meaning that if English sequences are to be interpreted as Japanese words, first language (L1) Japanese speakers have to adapt them to the stricter syllable structure of Japanese. Further, English is generally considered a stress-timed language (e.g., Pike, 1945; Bolinger, 1965) (although dialectal variation occurs, for example, with Jamaican English being described as more syllable-timed Wassink, 2001). It should be noted that the division of languages into stress-timed, syllable-timed and mora-timed has been questioned (e.g., Arvaniti, 2009) and there is no clear consensus on what phonetic measurements can clearly represent these proposed distinctions. However, this topic continues to be discussed and investigated from the perspectives of both

production and perception, taking into account such variables as speaking rate, syllable structure and phrasal prominence (e.g., Turk and Shattuck-Hufnagel, 2013; Fuchs, 2016). Assuming the difference in rhythm between English and Japanese is valid, speakers of a mora-timed language such as Japanese may parse the syllables of English differently from how English speakers would.

THE ENGLISH AND JAPANESE SOUND INVENTORIES

The General American English vowel inventory comprises 14 to 15 distinctive vowels, that is, /i:/, /ɪ/, /e/, /æ:/, /ɑ:/, /ʌ/, /ɔ:/, /ʊ/, and /u:/, including the rhotic vowel /ɜ:/, the mid-central vowel /ə/, and the diphthongs /eɪ/, /oʊ/, /aɪ/, /aʊ/, and /ɔɪ/ (Ladefoged, 2005). It should be noted though that in many American English varieties (e.g., Californian English, see Hagiwara, 1997) the distinction between /ɑ:/ and /ɔ:/ is neutralized in the direction of [ɒ:] (Nishi et al., 2008). In contrast to American English, the modern Tokyo Japanese vowel inventory includes five distinctive monophthongal vowels only, that is /i/, /e/, /a/, /o/, and /u/ (e.g., Ladefoged and Johnson, 2014; Kubozono, 2015b; see **Figure 1**), which are differentiated according to vowel height (high, mid, low) and tongue backness (front vs. back). Corresponding to the five short monophthongs, Japanese has five long monophthongs, i.e., /i:/, /e:/, /a:/, /o:/, and /u:/ (Kubozono, 2015b, p. 3).

Table 1 displays the Japanese and American English consonant inventory, respectively. Both inventories share the plosive consonants /p b t d k g/, the nasals /m n ŋ/, the approximants /j/ and /w/, as well as the alveolar fricatives /s/ and /z/. The English inventory has additional fricative consonants which do not exist in Japanese, namely /f v θ ð ʃ ʒ/. Japanese, by contrast, includes the bilabial fricative /ɸ/, as in “funē” (船, Eng. “boat”), and the alveolo-palatal fricatives /ç/ and /ʝ/, as in “shimbun” (新聞, Eng. “newspaper”) and “jagaimo” (じゃがいも, Eng. “potato”), respectively. Unlike English, Japanese does not have a contrast between /l/ and /r/, and there is no consonant in Japanese which exactly corresponds to either English /l/ or English /r/ (see Guion et al., 2000; Ohata, 2004). According to Guion et al. (2000, p. 2711), phonologically, the Japanese liquid consonant /r/ “might be considered similar to both English /l/ and /r/ [sic!],” but is typically produced as an alveolar or postalveolar apical tap [ɾ] (e.g., Vance, 1987; Labrune, 2012).

Unlike English which has two affricates only, i.e., /tʃ/ and /dʒ/, both produced at a postalveolar place of articulation, the Japanese inventory comprises four affricate consonants, namely the voiceless and voiced alveolar affricates /ts/ (“tsume”, Eng. “nail”) and /dz/ (“zurui”, Eng. “sly”), and the alveolo-palatal affricates /tʃ/ (“chuugoku”, Eng. “China”) and /dʒ/ (“jikan”, Eng. “time”).

While the English nasal inventory includes two consonants only, i.e., /n/ and /ŋ/, Japanese has a larger nasal inventory resulting from allophonic variation of the moraic nasal /N/, which does not have a specified place of articulation (e.g., Otake et al., 1996; Labrune, 2012; Kubozono, 2015a). For instance, /N/ is typically realized as the uvular consonant

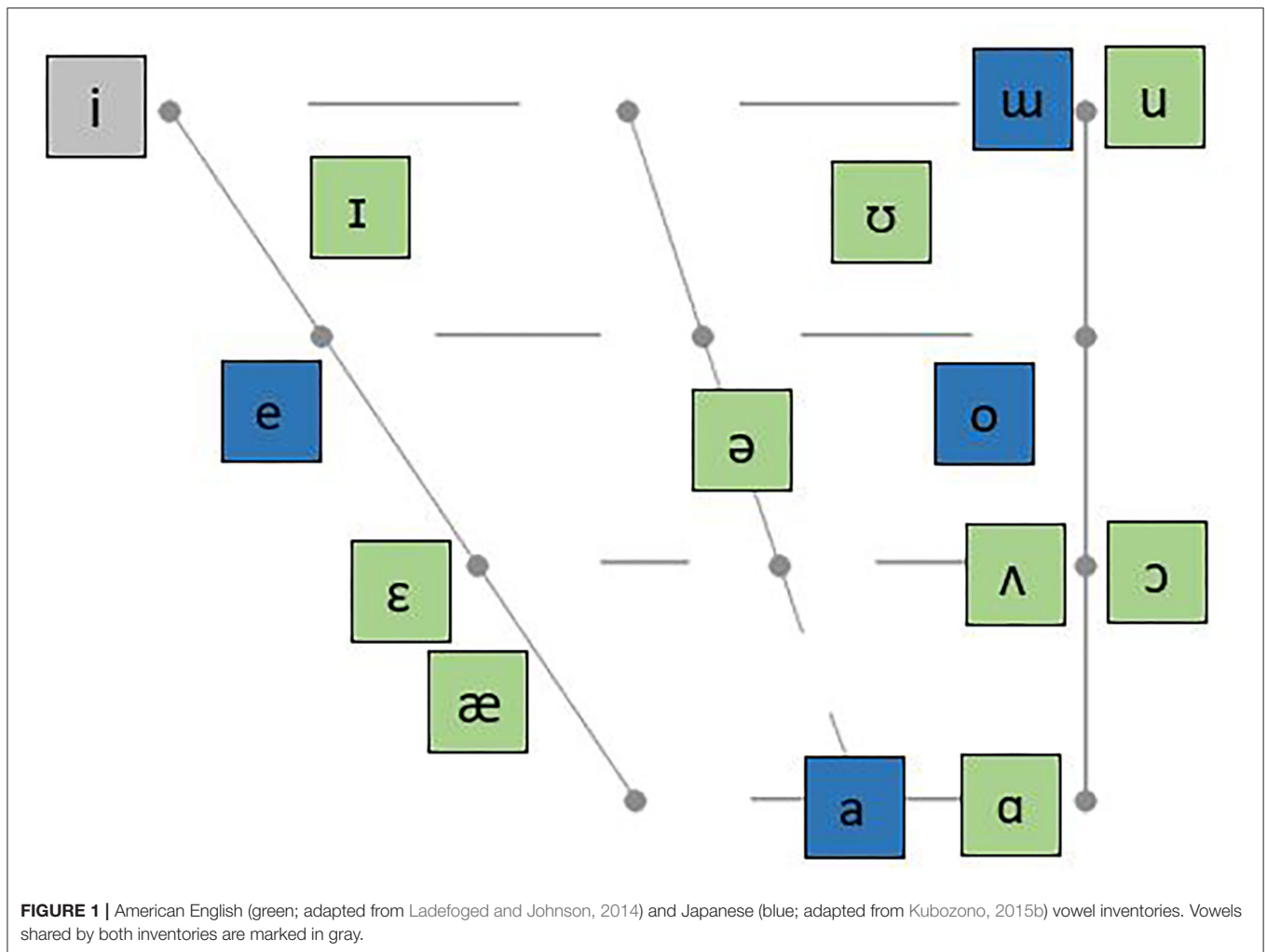


TABLE 1 | English (green) and Japanese (blue) consonant inventories.

	Bilabial	Labiodental	Dental	Alveolar	Alveolo-palatal	Postalveolar	Palatal	Velar	Uvular	Glottal	
Plosive	p	b	t	d				k	g	ʔ	
Nasal		m			n			ŋ	N		
Affricate				ts	dz	tʃ	dʒ				
Tap or Flap				r							
Fricative	ɸ	f	v	θ	ð	s	z	c	z	ʃ	ʒ
Approximant				j						j	
Lateral approximant				l				w			
Moraic consonants										/Q/, /N/	

Consonants shared by both inventories are marked in gray.

[N] particularly in slow speech. When occurring before the stops /p b m/, /N/ is realized as the bilabial nasal [m], when preceding the alveolar consonants /t d n/, the moraic nasal is produced as [n], and when occurring before the velar plosives /k g/, it is realized as [ŋ] (see Labrune, 2012, p. 133-4, for a detailed overview).

MONDEGREENS, SORAMIMI, AND SPEECH PERCEPTION

Mondegreens are often described as phenomena of auditory illusion (Kentner, 2015), and refer to listeners’ misperceptions of words or phrases in songs presented in the listeners’ native

language (“within-language misperceptions”, Beck et al., 2014, p. 2). However, misperceptions of song lyrics are not restricted to songs in the listener’s native language, but might also occur across different languages, which Kentner (2015, p. 1) refers to as “cross-linguistic mondegreens”. This means that listeners may perceive foreign lyrics, including both individual words and complete phrases (Otake, 2007), through their native language and adapt the foreign stimuli to their native language system. Typically, “mondegreening”, resulting from either within-language or across-language misperceptions, leads to the creation of new, usually amusing, meanings (e.g., Beck et al., 2014). The term “Mondegreen” goes back to Wright (1954) who described her mishearing of the original line “layd him on the green” in the seventeenth-century Scottish folk song “The Bonny Early O’ Murray” (see Olson, 1997) as “Lady Mondegreen”. In Japanese, mondegreens are known as *soramimi* (空耳, “mishearing”), which has lent its name to the popular Japanese TV show *Soramimi Hour* (*soramimi awā*). For the sake of entertainment, Japanese listeners are invited to submit misheard English song lyrics to the show which are then presented in short comical video clips to the audience and which are rated and awarded by the show hosts based on the degree to which they share the mishearing.

From a linguistic point of view, mondegreens and *soramimi* can be considered a “valuable tool to induce plasticity within the auditory system” (Beck et al., 2014, p. 2), but only few studies so far have systematically examined the actual processes underlying the phenomenon of misperceiving song lyrics and adapting foreign lyrics to the listener’s native language system. To the best of our knowledge, to date there is only one study (Otake, 2007) that investigated phonological processes involved in listeners’ misperceptions of foreign song lyrics: Otake (2007) examined a corpus of 194 English-Japanese *soramimi* containing interlingual near-homophonic words and phrases, i.e., English words and phrases which were misperceived as near-homophonic Japanese targets by Japanese listeners (e.g., the English source lyric “psycho” was perceived as Japanese “saikō” [saikou], “the highest”, see Otake, 2007, p. 778). Otake identified different reoccurring phonological processes, including segment omissions, sound substitutions and additions (see Sections Integration Processes for Foreign Language Material in Japanese and Methodology, this paper), underlying the Japanese adaptations of English auditory input identified in the *soramimi* samples. As Otake (2007, p. 779) concludes, the *soramimi* examples revealed “the same kinds of processes as appear in adaptation of foreign loan-words in Japanese” (see also Peperkamp et al., 2008), showing that Japanese listeners may perceive English words and phrases through the phonological filter of their native language.

In order to give further insights into the phenomenon of cross-linguistic auditory misperception and the phonological and morphological processes involved, the present study set out to analyze 60 examples of English-Japanese *soramimi* taken from the above-mentioned Japanese TV show *Soramimi Hour* (see Section Methodology, for a detailed description of the speech corpus).

CUES TO PROCESSING CONNECTED SPEECH

When processing spoken language, listeners mainly use the acoustic signal, but may also use visual cues (McGurk and MacDonald, 1976) and even tactile cues, when available (Gick and Derrick, 2009). The McGurk Effect has been shown to occur among adult listeners, whereby speech processing can be affected by visual cues—in this study, silent videos of speakers producing stop consonants—that do not match the auditory signal (McGurk and MacDonald, 1976). Processing spoken language is a challenge which contends with factors that affect production, including coarticulation, emotional effects, interspeaker variability, accentedness, speaking rate, and background noise (e.g., Ladefoged and Broadbent, 1957; Mann and Repp, 1980; Brent et al., 2013; Dame et al., 2013). Therefore, listeners use as many cues as possible in order to avoid mishearing and misunderstanding the signal. Using these cues are examples of bottom-up processing, but listeners also use knowledge about the structure and patterns of the language in order to interpret the signal, a process known as top-down processing (e.g., Warren, 1970; Ganong, 1980). For example, if a person listening to English hears “the”, they know that the following word is likely to be a noun or an adjective. It has been shown that a sound that is at the acoustic boundary between two phonemes will be interpreted as one phoneme or the other based on whether the resulting word would form a real word (Ganong, 1980). For example, if a sound is at the acoustic boundary between English /d/ and /t/, if it is presented as the onset to “ask”, listeners will tend to hear “task” rather than “dask”, because only the former is a real word. As such, speech processing is influenced by knowledge of the language.

Especially in a noisy environment where the acoustic signal is degraded, listeners try to use as many cues as possible to interpret speech. Top-down processing, therefore, is likely to be relied on more when the acoustic signal is noisy. As such, due to the combined effect of music in the signal, song lyrics are potential triggers for mishearing lyrics, also known as mondegreens (see Section Mondegreens, *Soramimi*, and Speech Perception). Balmer (2007) proposes the Phonetic Ambiguity Hypothesis which states that the phonetic ambiguity present in song lyrics is what triggers mondegreens. Also, in listening to songs, listeners process the words and melodies interactively (Gordon et al., 2010; Tang, 2015), and further, listeners do not have the contextual cues that would be available in conversational speech (Tang, 2015). Furthermore, due to the less clear signal in music, lyrics are ideal targets for manipulation when paired with visual cues and written words. Research has found that phonological as well as orthographic factors can have a priming effect in picture-word interference tasks (Chéreau et al., 2007; Damian and Bowers, 2009). Such an effect may extend to *soramimi*, which are presented as visual scenes as well as Japanese words highlighting the misheard lyrics. The presentation of these multiple cues is likely to prime the intended interpretation for viewers. It should also be noted that previous research (e.g., Starr and Shih, 2017) has suggested that Japanese speech produced in

TABLE 2 | Example *soramimi* (No. 190), corresponding IPA transcriptions and phonological processes identified in this example.

	Lyrics	IPA	Processes
English original	<i>I knocked on every door</i>	[ana ^k ta/ne/ ^v ido:]	4 SUB 1 INS 1 DEL
Japanese misperception	<i>anata nee buri dou</i>	[anata/nee/bu ^β ri do:]	1 BTR 8 NoCh

Phonological processes (SUB, INS, DEL) are color-coded.

songs may be different from spoken speech with regard to certain aspects such as phonology; the fact that the input analyzed in the present study consists of sung language may therefore also have an impact on native Japanese speakers' perception thereof.

A last aspect to consider when it comes to processing connected speech are potential influences of social information. Strand (1999), for example, found that gender stereotypes maintained by listeners significantly alter their perception along the /s/-/ʃ/-continuum. Similarly, Johnson et al. (1999) observed that speaker gender influenced how listeners located the phoneme boundary between the English vowel targets /u/ and /ʊ/. With regard to the (mis-)perception of song lyrics, as analyzed in the present study, listeners' perceptions of individual sounds and/or words might therefore—at least to some extent—be affected by the gender (and potentially other social information) of the singer.

METHODOLOGY

In order to examine which phonological and morphological processes are involved in the misperception of English song lyrics by Japanese listeners, a speech corpus of 60 English-Japanese cross-linguistic mondegreens⁵ (*soramimi*) was analyzed. This corpus was assembled based on a large pool of “felicitous” *soramimi*⁶ from various episodes of the Japanese TV-show *Soramimi Hour*, broadcast since 1992, from which 60 examples were then randomly selected. After downloading the samples from the online platform YouTube, both the original English lyrics (source lyrics) and the Japanese misperceptions were transcribed into IPA by two trained phoneticians based on the extracts presented in the clips (see an example in **Table 2**). The individual samples represent short phrases from different English song lyrics and the corresponding Japanese mishearings, which differed in syllable number (English: $M = 6.78$, $SD = 2.39$; Japanese: $M = 7.22$, $SD = 2.55$), depending on where the misperception occurred. Based on a comparison between the English source lyrics and the Japanese misperceptions, the adaptations and changes identified in the Japanese samples were categorized according to one of three phonological processes, i.e.,

⁵We use the term *English-Japanese mondegreens* to refer to English song lyrics which have been adapted to the Japanese language by native Japanese listeners.

⁶As “felicitous” we labeled such *soramimi* to which the show's hosts gave high ratings and reacted positively and where therefore perception of the mishearing was arguably shared by them.

sound substitution (SUB), vowel insertion (INS), and segment deletion (DEL) (see Otake, 2007; see **Table 2**). As outlined in Section Integration Processes for Foreign Language Material in Japanese, these processes are most frequently observed when it comes to adapting English loanwords to the Japanese language (e.g., Loveday, 1996; Nian and Jubilado, 2011; Scherling, 2013; Shoji and Shoji, 2014). Sound substitutions are likely to occur when English vowels (e.g., /ɜ/, /æ/) and consonants (e.g., /θ/, /ð/, /r/, /l/, /v/) which are not part of the Japanese inventory are encountered and are replaced by phonologically similar Japanese sounds (see e.g., Stanlaw, 2004; Pintér, 2015). Alongside replacing an English sound with a Japanese target, an epenthetic vowel—typically /i/, /o/, or /u—might be inserted in an English source word to break up consonant clusters and maintain open syllables, given that both consonant clusters and closed syllables (with the exception of syllable-final /n/) are not permitted in Japanese (e.g., Dupoux et al., 1999). When it comes to segment deletion, despite being less frequently observed than vowel insertion (see Shoji and Shoji, 2014), English inflectional suffixes or word-final consonants might be deleted to avoid closed syllables.

Alongside the three categories (SUB, INS, DEL) outlined above, an additional morpho-phonological category was introduced for boundary transgressions (BTR). This category contained Japanese adaptations resulting from a reanalysis of the morphological boundaries of an English phrase and thus subjecting it to Japanese morphology. For example, the English phrase [saɪts ɔn ju] (“sights on you”) was perceived as the Japanese word [saɪsɔ:nʲu:] (saisōnyū, Eng. “reinsertion”), involving a reanalysis of the word-final consonant /s/ in English [saɪts] as the onset of the second syllable in the Japanese misperception.

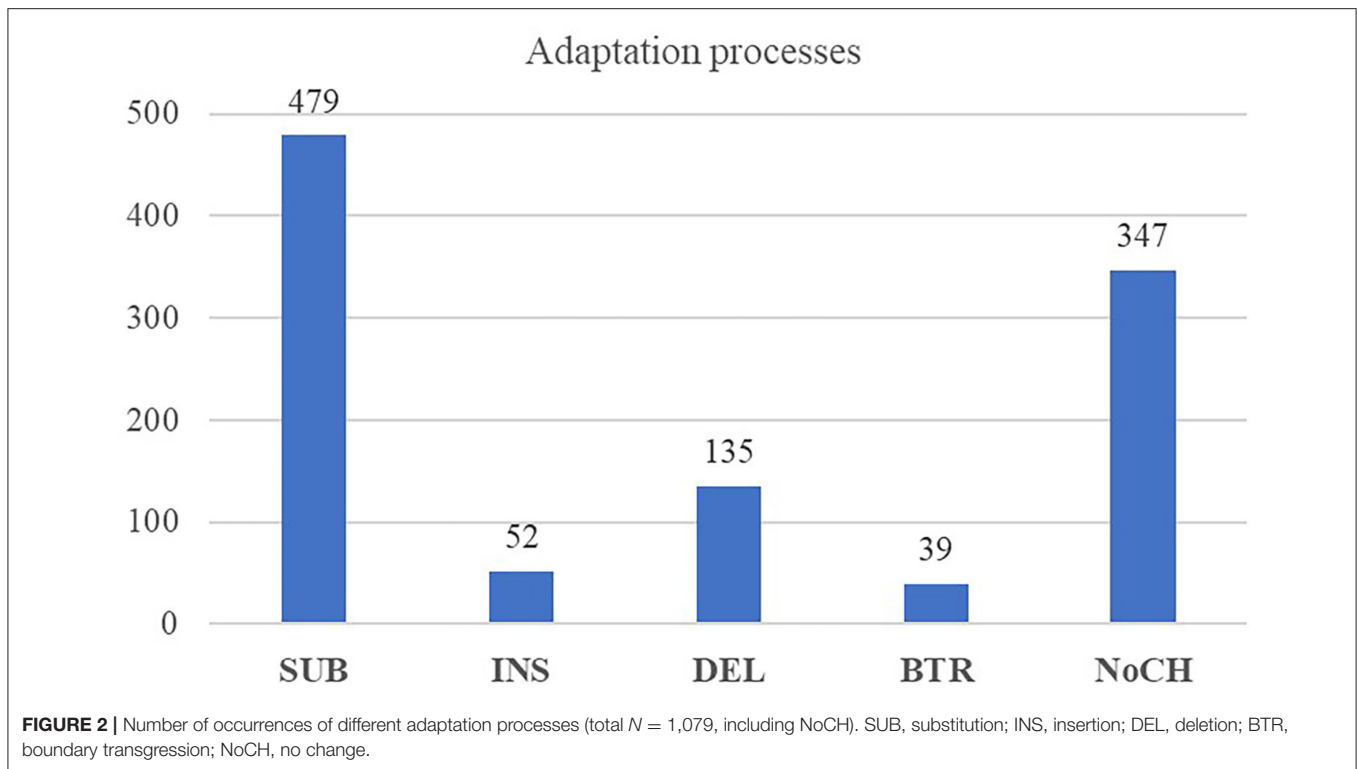
A fifth category labeled “no change” (NoCh) was included for sounds or sound sequences which did not reveal any phonological or morphological differences between the source lyrics and the misperceived lyrics. **Table 2** gives an example of a *soramimi* taken from our corpus, including the English source lyrics, the Japanese misperceived lyrics and the corresponding IPA transcriptions. The phonological processes identified in this example are color-coded.

ANALYSIS

As displayed in **Figure 2**, out of the total number of phonological adaptation processes identified in the corpus ($N = 1,079$), the largest number is constituted by the process of sound substitution ($N = 479$), i.e., by the replacement of English sounds by sounds that are part of Japanese phonology.

Oftentimes, SUBs affected vowels or consonants that are part of the English sound inventory, but do not exist in the Japanese inventory, while in some other cases sounds that do exist in Japanese were replaced by other Japanese sounds. For example, the English phrase [hoʊp kæn kip mi tʊgeðə] (“Hope can keep me together”) was misheard as [hokenkiN mitsuketa] (“Hokenkin mitsuketa”—“I found the insurance money”). In this mishearing, the following substitutions can be observed:

Vowels (Eng. → Jpn.): oʊ → o / a → e / u → ɯ / ε → e / ə → a



Consonants (Eng. → Jpn.): $n \rightarrow \eta$ / $p \rightarrow N$ / $t \rightarrow ts$ / $g \rightarrow k$ / $\delta \rightarrow t$

In the case of the vowels, there are four that do not exist in Japanese, namely /ou/, /æ/, /ɛ/ and /ə/, all of which were replaced by vowels that are the closest approximations in the Japanese vowel inventory. The vowel /u/ itself is not part of the Japanese vowel inventory, but a similar sound, [u], is, which is why it substitutes for /u/. As for the consonants, while /n/ is also phonemic in Japanese, when positioned in a syllable coda it is realized as a moraic nasal /N/ and is known to assimilate to the following sound (Vance, 2008) as in the example above. This is also what happens in the case of /p/, where the replacement with Japanese /N/ occurs because it precedes a bilabial nasal /m/ and thus becomes subject to nasalization. The plosive /t/, again, exists in Japanese, but is restricted with regard to the vowels it can combine with. While the syllables /ta/, /te/, and /to/ are licensed, /tu/ and /ti/ are not. For these two cases, /t/ is replaced by Japanese affricates, i.e., /ts/ combines with /u/, and /tʃ/ with /i/. As a consequence, English /tu/ is replaced in this mishearing with /tʃu/. Finally, the English fricative /ð/ is another unknown consonant in Japanese, and is thus substituted by the alveolar stop /t/, which is a common process in languages that do not contain interdental fricatives (Blevins, 2006).

Another example is the mishearing of [wəgʊr ɪz əkɪs] (“What good is a kiss”) as [ʋəkare zakessʋ^β] (“Wakaresakessu”—“It’s a goodbye drink”). In this example, the following substitutions can be identified:

Vowels (Eng. → Jpn.): $\text{ə} \rightarrow \text{a} (2x)$ / $\text{ɪ} \rightarrow \text{e} (2x)$

Consonants (Eng. → Jpn.): $w \rightarrow \text{ɰ} / g \rightarrow k$

As in the previous example, the English schwa sound /ə/ is replaced by /a/ in the Japanese mishearing, as is the English close central rounded vowel /ɪ/. While there is only a superficial phonetic similarity between /ɰ/ and /a/ in that they are both unrounded, it can be argued that this misperception is construed to render this sequence meaningful. This means that, in this particular case, it may be that top-down processing, the impetus to make and perceive a real word, overrides the lack of similarity between the original vowel and the perceived vowel. The vowel /ɪ/, in this example, is twice perceived as an /e/-vowel by the Japanese listener, even though there is a more similar vowel, namely /i/, available in the Japanese inventory. However, it has to be noted that the vowel /i/ in the original English lyrics, especially for /ɪz/, is rather centralized and indeed could be considered to be somewhere between /ɪ/ and /ə/, which might explain the substitution. The two consonant changes are essentially substituting the non-phonemic sound /w/ with its closest approximation in Japanese, i.e., /ɰ/, and changing the voicing of the velar stop (/g/ to /k/), possibly because the background music makes it difficult for listeners to perceive whether the stop is voiced or voiceless, so place and manner of articulation in the previous example are retained while voicing changes in the Japanese misperception. In addition, stops in Japanese, according to Fujimoto, are “more aspirated word-initially than word-medially” (Fujimoto, 2015, p. 201) and thus the distinction between voiced and voiceless stops may be more difficult to draw in medial positions, as in this case the velar stop /k/ in “wakare”.

As a last example for substitutions, the following mishearing will be discussed: [dɒntʃɑ nɪd səm lʌv tənɑtʃ] (“Don’t you need some love tonight”), which was perceived as [tɔ:tʃɑN nʲi:tʃɑN jotte nai] (“Tōchan, niichan yottenai”—“My dad and older brother are not drunk”). The following sound substitutions can be observed in this example:

Vowels (Eng. → Jpn.): ə → a / ʌ → o / ə → e

Consonants (Eng. → Jpn.): d → t / tʃ → tʃ / ds → tʃ / m → N / l → j

Once more, it can be observed that the English schwa sound, not being part of the Japanese vowel system, was perceived as /a/ and /e/, respectively, by the Japanese listener. The corpus data here shows that /a/ is the substitute in the large majority of cases, except for those where the schwa is preceded by /t/ or /θ/. The vowel /ʌ/ in “love” is replaced by the closest vowel in the Japanese vowel inventory, which is the vowel /o/. For the consonants, word initial /d/ is substituted by its voiceless counterpart /t/, possibly so perceived because of the increased aspiration in /d/ in the sung performance. There are two substitutions with the Japanese affricate /tʃ/: In the case of English /tʃ/, the Japanese affricate /tʃ/ is the most similar sound and thus substitutes for the English affricate. On the other hand, in the case of English /ds/, arguably through voicing assimilation, the /d/ might be more similar to /t/. In Japanese, this would create the illicit sound sequence /tsə/ or—after substitution of /ə/ by /a/—/tsa/; hence, it is substituted with the closest Japanese sound sequence /tʃɑ/. Finally, an interesting aspect is the omission of the /n/ sound in /dɒntʃɑ/, which seems to reappear later, at the end of the sequence in /tɔ:tʃɑN/, which we would argue indicates a perseverative effect of the nasal.

While a considerable amount of sound substitutions was identified in our corpus, insertions were comparatively rare ($N = 52$). This seems reasonable as the addition of non-existent vowels would alter the rhythm of the song lyrics and would therefore make it less likely to be perceptually shared by others. Consequently, most of the insertions identified in our corpus relate to those necessary in Japanese phonotactics, i.e., inserting vowels to break up consonant clusters. In the following, we will discuss three examples of such insertions.

Example 1:

Eng. [nɑ:θɪn kɑn haʊld ʌs ɑn:ɑθɪn kɑn kɪp ʌsdɑwn] (“Nothing can hold us and nothing can keep us down”) → Jpn. [nadze ka hon dʌʃi:ta, nadze ka ʃitto ʃi:ta] (“Naze ka hon dashita, naze ka hitto shita”—“For some reason, I published a book and for some reason, it became a hit”)

Example 2:

Eng. [kɪdz kʊdn hæ dʒæk ðeɪ tɹaɪ tɹaɪ tɹaɪ] (“Kids couldn’t hurt Jack. They tried and tried and tried”) → Jpn. [k i tsuβkuβ ΦuβN dʒa itai itai itai] (“Kitsuku hunja itai itai itai”—“It really hurts when you step on it”)

Example 3:

Eng. [(ð)ə naɪ deɪ dɹoʊv owl dɪksi dɑwn] (“The night they drove old Dixie down”) → Jpn. [inaɪ dze dorobo: dekʲi:ci da] (“Inaize dorobo. Dekishi da.”—“The robber’s gone. He drowned”)

In Example 1, we can see that the vowel /i/ was inserted twice. While in the second case, it is inserted to break up what would otherwise be an illicit consonant cluster (/ds/), in the first instance, the situation is more complex as the original lyrics already include a licit sound combination, namely /sa/. Since the case for insertion to preserve Japanese phonotactics cannot be made here, it may be argued that the mishearing up to this point (“hon das”) primes the listener to mentally complete what is an established idiom in Japanese (“hon dashita”—“published a book”), which would be another example of top-down processing.

In Example 2, the epenthetic vowel [uβ] is used to break up the English consonant cluster /dzk/. Japanese /N/, as we have argued before, may have been perceived and thereby inserted, due to the perseverative effect of the—otherwise deleted—/n/ in the word /kʊdn/, i.e., the sound appears later in the sequence in the mishearing by way of metathesis.

Example 3 is another case where consonant clusters are altered to make the sequence adhere to Japanese phonotactics. The original lyrics include two clusters, /dr/ in “drown” and /ks/ in “Dixie”, which are modified in the mishearing by inserting the epenthetic vowel /o/ after /d/, and /i/ after /k/.

Out of the 135 deletions identified in the corpus, $N = 99$ were consonant deletions, showing that consonants were far more likely to be deleted than vowel targets. In $N = 15$ out of these deletions, word-final, mainly unreleased English plosives were omitted in the Japanese misperceptions, as shown in the following examples:

Example 4:

Eng. [bæk ɪn hæʊl weə ðeɪ gɑtʃ] (“Back in the hole where they got”) → Jpn. [gakʲi no ho: ɥɑ dekaɪ] (“gaki no hō wa dekaɪ”—“the little one’s is bigger”)

Example 5:

Eng. [dɒntʃɑnɪd səm lʌv tənɑtʃ] (“Don’t you need some love tonight”) → Jpn. [tɔ:tʃɑN nʲi:tʃɑN jotte nai] (“Tochan, niichan yottenai”—“My dad and older brother are not drunk”)

Example 6:

Eng. [hoʊpʃ kɑn kɪp mi tuɡeðə] (“Hope can keep me together”) → Jpn. [hokeŋkʲiN mʲitsuβketa] (“Hokenkin mitsuketa”—“I found the insurance money”)

In these examples, the post-vocalic alveolar plosive [tʃ], lacking an audible release burst in the English syllables [gɑtʃ] and [nɑtʃ], and the unreleased word-final English bilabial plosive [pʃ] in [hoʊpʃ] were deleted in the corresponding Japanese misperceptions. A possible phonetic explanation for this is the observation that syllable-final plosive consonants which are produced without an audible release are usually less intelligible than released plosives (e.g., Lisker, 1999; Nozawa and Cheon, 2014). This is particularly the case if the plosive target is preceded by a diphthong, as in Examples 5 and 6 above, potentially resulting from within-vowel formant movements which may impede listeners’ ability to disentangle consonant and vowel components (see Lisker, 1999). Alongside phonetic considerations, deleting unreleased final plosives is likely to be a consequence of subjecting English input to Japanese phonotactic

constraints. In Japanese, the plosives /p b t d k g/ exclusively occur in syllable-initial position given that closed syllables are disallowed (e.g., Ito and Mester, 2015). Hence, it can be argued that deleting syllable-final English consonants—whether they are released or not—aims to maintain open syllables, as shown in the above examples. This is, of course, not restricted to final plosive consonants, but may also affect other English consonants occurring in syllable-final position and thus constituting closed syllables. For instance, the final fricative /v/ in the English lyrics [aɪsɛdawkʊ:t hi ɛvəli:v] (“I said ‘How could he ever leave’”) was deleted in the Japanese misperception [asedakʊ^β ɕiabʊ^βɾʲi] (“Ase daku hi aburi”—“Sweating and burning”) to maintain an open syllable.

Similar observations were made with regard to English consonant clusters where the second consonant of the cluster was omitted in the Japanese misperception. In Example 2 above, the word [tʁai] in the English lyrics [kɪdz kʊdn hɛdʒæk ðeɪ tʁai tʁai tʁai] (“The kids couldn’t hurt Jack, they tried, tried, tried”) was reinterpreted as [itai] (“it hurts”) in the Japanese misperception [k i tsʊ^βku^β Φʊ^βN dʒa itai itai itai], involving a deletion of the English approximant [ɹ] and thus avoiding the clustering of two successive consonants in Japanese.

In terms of segment deletions, it was further observed that diphthongs occurring in the English source lyrics were frequently perceived as monophthongs in Japanese:

Example 7:

Eng. [ʔɪdoʊn meɪkɪʔizeɪ] (“Ya don’t make it easy”) → Jpn. [ittɕo:me de i: dze] (“Icchōme de ii ze”—“You can drop me off after just one block”)

Example 8:

Eng. [meɪ teja gawd ɔmaterɪ goʊna kʊtʃudɑ] (“[Let] me tell you God Almighty gonna cut you down”) → Jpn. [mʲi te jo, kore marʲiko no ku^βtsu^β da] (“Mite yo kore Mariko no kutsu da”—“Look, it’s Mariko’s shoes”)

In these examples, the English diphthongs /oʊ/, /eɪ/, and /eɪ/ were reinterpreted as the Japanese monophthongs /o/, /e/, and /i/, respectively, in each case involving the deletion of one of the two adjacent vowel targets in the diphthongs. As discussed in Section Integration Processes for Foreign Language Material in Japanese, the Japanese vowel inventory does not generally contain diphthongs (with a few exceptions only, see e.g., Kubozono, 2005) and, hence, native Japanese listeners might encounter difficulties perceiving English diphthongs and replace them with monophthongs instead.

As shown in **Figure 2**, the lowest number of processes identified in the speech corpus is constituted by boundary transgressions ($N = 39$), that is, reanalyzing morphological boundaries of an English phrase to adapt it to Japanese morphology. As illustrated in the following examples, boundary transgressions were identified in cases where an English consonant in coda-position was perceived as the syllable-onset of the following word in Japanese:

Example 9:

Eng. [jʌ:downowmane:] (“You don’t know my name”) → Jpn. [jɔʊ^β dou^β nomanee] (“Yo! Do! Nomanee!”—“Hey, how about a drink?”)

Example 10:

Eng. [sarts ɔn ju] (“sights on you”) → Jpn. [sai so:nʲʊ^β:] (“saisōnyū”—“reinsertion”)

Example 11:

Eng. [ɛvɪdaɪkɛskaj] (“Every darkest sky”) → Jpn. [ɛbʲi dake ssʊ^βka] (“Ebi dake ssuka”—“Only shrimp?”)

In Example 9, the coda consonant in English [down] was perceived as the onset consonant in the Japanese word [nomanee]. Similarly, in Examples 10 and 11, syllable-final English /s/ in [sarts] and [daɪkɛs] was reanalyzed as the onset consonant in the Japanese targets [sai so:nʲʊ^β:] and [ssʊ^βka], respectively. Example 10 additionally shows a deletion of the English alveolar plosive /t/ preceding the coda consonant /s/, which allows for the open syllable [sai] in the Japanese misperception.

DISCUSSION AND CONCLUSION

As our analysis has shown, the processes that underlie the majority of perceptual changes between the original English song lyrics and their Japanese misperceptions largely follow the patterns also observable in the perception of English in general and of English loanwords in particular (e.g., Loveday, 1996; Stanlaw, 2004). This confirms our initial hypothesis that the perception of foreign language input is likely to be filtered through and affected by L1 phonotactics and hence subject to the same processes as loanword integration *via* the auditory channel.

In essence, the reinterpretation of English auditory input occurs through the phonological filter of Japanese, which is reflected in the frequent occurrence of various phonological and morphophonological processes, namely sound substitutions, insertions, segment deletions, and boundary transgressions. All these processes are aimed at making the perceived input adhere to Japanese phonotactics by maintaining open syllables, breaking up consonant clusters by means of epenthetic vowels as well as by reanalyzing word-final consonant syllable codas as syllable onsets of the following phrase (see e.g., Dupoux et al., 1999; Kubozono, 2002).

The bulk of evidence for perceptual assimilation is reflected in sound substitutions, which by far make up the largest number of phonological processes identified in our corpus. As discussed earlier in this paper, perceptual assimilation, as defined by Peperkamp et al. (2008, p. 131), describes “a process that applies during speech perception and that maps non-native sound structures onto the phonetically closest native ones”. The examples in our corpus have shown that English sounds that are non-existent in Japanese were systematically, though not exclusively, replaced by the closest Japanese sounds. This

occurred not only for individual segments that are not part of the Japanese sound inventory, such as the dental fricatives /θ/ and /ð/ or the lateral approximants /r/ and /l/, but also for sound combinations that are not licit in Japanese, such as consonant clusters or certain CV-combinations, for example English /tu/ or /si/ (see e.g., Dupoux et al., 1999). In some particular cases, we identified English sounds being replaced by non-similar Japanese sounds, such as [ɰ] being substituted by [a]. For these cases, we argue that they might represent examples of top-down processing, which, as outlined in Section Cues to Processing Connected Speech, refers to listeners' ability to create meaningful linguistic units out of unclear auditory input (e.g., Warren, 1970; Ganong, 1980). When it comes to the perception of English song lyrics, this means that the listeners' instinct to perceive a meaningful word or phrase through their own language filter is so strong that this overrides the lack of phonological similarity between the original and the perceived segment. This could be seen in cases where the expectation of a certain word or phrase was already primed, thus arguably leading the listener to falsely perceive a sound despite a complete absence of similarity between the original English and the misperceived Japanese sound. In other cases, we found that voiced stops occurring in initial position in the English original would be substituted by voiceless stops occurring in medial position in the Japanese misperception, which we attributed to a lack of aspiration in stops in medial position in Japanese (see e.g., Fujimoto, 2015), making the voicing distinction more difficult to draw. Additionally, it might be argued that the misperception of plosive voicing results from adverse listening conditions, such as a high level of background noise, which has previously been shown to affect the accurate perception of English plosives (Brent et al., 2013).

As for insertions, in line with what we expected, these were less frequently observed. The reason for this, we argue, is that including additional vowels would not only extend the phrase in terms of syllable number, but would also thereby break up the original rhythm of the song, making the reinterpretation less likely to be perceived, and/or shared by others. Where insertions did occur, this was mostly to break up consonant clusters in the English original, i.e., to make the mishearing adhere to Japanese phonotactics. In other, rarer cases, additional syllables were inserted to create a meaningful utterance or complete an idiom that was perceived, priming the listener to it, but where the final syllable of that idiom was not part of the original lyrics. Finally, an interesting phenomenon we identified was the occasional insertion of /N/ which occurred at a different position in the original. The reason, we argue, why this nasal was perceived at a later position in the mishearing by way of metathesis is due to both a perseverative effect of the nasal and due to top-down processing, in which the expectancy to hear a meaningful word led to the reassignment of the sound to a different position.

Deletions served a similar purpose to insertions, namely to maintain open syllables, by deleting sounds such as coda consonants. In many cases, these consonants were unreleased English voiceless plosives which are less easily perceivable than plosives produced with an audible release (e.g., Lisker, 1999). In other cases, deletion was applied to consonant clusters as an alternative to epenthetic vowel insertion which, as discussed

above, would have interfered with the rhythm of the original lyrics. Additionally, the input being aural, final consonant deletion was an anticipated outcome, as unvoiced consonant endings may lead to a "lack of perception" (Shoji and Shoji, 2014, p. 9).

The last process, boundary transgression, was the most systematic in the sense that it exclusively occurred when an English word ended with a consonant coda, which was then—instead of omitting it—reanalyzed as the onset of the subsequent word or phrase in the Japanese mishearing. In some cases, this was facilitated by the following phrase having a vowel onset while in other cases, boundary transgressions combined with insertions of epenthetic vowels to ensure licit Japanese sound combinations.

As the above discussion shows, the last three processes—insertion, deletion, and boundary transgression—serve a common purpose, namely to maintain the legality of sound combinations in the misperception by either adding a sound, eliding a sound, or shifting the morphological boundaries of the original lexeme to ensure phonotactic compatibility. Simultaneously, substitution works to adapt the input to the Japanese sound inventory. As such, these four processes condition, affect and trigger each other in complex ways and therefore effectively reflect the impact of top-down processing and perceiving foreign language input through the listeners' L1 filter. We argue that the choice of the process to ensure adherence to Japanese phonotactics is clearly a factor of priming and expectancy bias, i.e., which word or phrase is anticipated, and that therefore, while seemingly random, whether a sound gets inserted, deleted or morphologically reanalyzed is a direct consequence of top-down processing.

While the findings obtained in the present analysis provide insights into the processes governing the misperception and reinterpretation of foreign language input, it needs to be taken into consideration that our speech corpus only included "felicitous" *soramimi*, that is, those that were immediately reproducible to the show hosts upon playing the clips. There are many other mishearings presented in the show which the hosts struggled with or found impossible to reproduce, hence it is not possible to generalize the present findings to include all *soramimi*, as some are clearly triggered by factors related to the listener and the listening situation. It is also possible that there was some editorial influence or that some of the examples were fabricated for the show; however, even if this did occur, the examples need to be similar enough to the English lyrics to be plausible to the audience. For the present analysis, we selected the best-practice examples in order to understand not merely the processes underlying crosslinguistic mondegreens, but also to get an idea of the conditions that need to be met for a mondegreen to be perceptually shared by other members of the speech community as well. While we cannot assume that our findings are generalizable to naturally-occurring data, the current analysis clearly shows patterns in terms of what processes are used and which ones are used more often.

Another potential limitation of a study of this kind, as addressed by Otake (2007), is that there is no information available concerning the listeners' linguistic background, that is,

whether they are, for instance, proficient speakers of English as a second language (L2) or (quasi-)monolingual Japanese speakers. This is of importance because native Japanese listeners who are highly proficient speakers of L2 English might be more likely to perceive the English lyrics accurately compared to less proficient speakers of L2 English (see e.g., Kaneko, 2006, for a discussion of the role of L2 proficiency on vowel perception).

Finally, the findings of this study are also limited in the sense that it is not possible to determine the specific environment in which the mishearing occurred and whether there were potential factors involved that facilitated the triggering of the misperception, such as music played in a club, on the radio, or in a car with multiple background noises (see Lecumberri et al., 2010, for a review of studies examining the effects of adverse listening conditions on the perception of non-native speech). Nevertheless, the current findings could be used as a starting point for further work involving controlled experiments with listeners of varying linguistic backgrounds as well as different listening contexts, in order to tease apart these influences.

As mentioned in Section Cues to Processing Connected Speech, priming through orthography and visuals might also play an important role in the misperceptions presented in *Soramimi Hour*. A potential follow-up study, therefore, could be to play English song lyrics used on *Soramimi Hour* to native Japanese listeners in various iterations: Only the music, with

subtitles, with images but without subtitles, then with images and subtitles, and to compare whether and to what extent each of these factors contributes to the success of the mishearing and hence, to what extent such primes “manipulate” the perception of the participants. This would help to determine the specific contributions of the different factors (lyrics, subtitles, images) to the inducing of a particular reinterpretation.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

JS: study design, transcription and analysis of data, and manuscript writing. LK: analysis of data, manuscript writing, and tables and figures. NK: transcription and analysis of data and contributed to, reviewed and edited the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

The authors acknowledge the financial support by the University of Graz.

REFERENCES

- Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica*, 66, 46–63. doi: 10.1159/000208930
- Balmer, F. (2007). *Phonetic, Phonological and Prosodic Triggers for Mondegreens*. Unpublished manuscript. Available online at: [http://www.ambiguetae.uni-tuebingen.de/minigrad/balmer/Balmer_\(2007\)_Phonetic,_Phonological_and_Prosodic_Triggers_for_MondegreensFINAL.pdf](http://www.ambiguetae.uni-tuebingen.de/minigrad/balmer/Balmer_(2007)_Phonetic,_Phonological_and_Prosodic_Triggers_for_MondegreensFINAL.pdf) (accessed October 20, 2021).
- Beck, C., Kardatzki, B., and Ethofer, T. (2014). Mondegreens and soramimi as a method to induce misperceptions of speech content: influence of familiarity, wittiness, and language competence. *PLoS ONE* 9, e84667. doi: 10.1371/journal.pone.0084667
- Blevins, J. (2006). New perspectives on english sound patterns: ‘natural’ and ‘unnatural’ in evolutionary phonology. *J. Engl. Linguist.* 34, 6–25. doi: 10.1177/0075424206287585
- Bolinger, D. (1965). “Pitch accent and sentence rhythm,” in *Forms of English: Accent, Morpheme, Order*, eds. T. Kanekiyo and I. Abe (Harvard: Harvard University Press), 139–80.
- Brent, J. A., Patra, H., and Vaz, P. C. (2013). Effects of noise and speaker’s language background on plosive perception. *J. Acoust. Soc. Am.* 134, 4229. doi: 10.1121/1.4831538
- Chéreau, C., Gaskell, M. G., and Dumay, N. (2007). Reading spoken words: orthographic effects in auditory priming. *Cognition* 102, 341–360. doi: 10.1016/j.cognition.2006.01.001
- Dame, M. A., Patra, H., and Vaz, P. C. (2013). Effects of speakers’ language background on speech perception in adults. *J. Acoust. Soc. Am.* 134, 4063. doi: 10.1121/1.4830833
- Damian, M. F., and Bowers, J. S. (2009). Assessing the role of orthography in speech perception and production: evidence from picture–word interference tasks. *Eur. J. Cogn. Psychol.* 21, 581–598. doi: 10.1080/09541440801896007
- Dower, J. W. (1999). *Embracing Defeat*. New York, NY: W. W. Norton and Company.
- Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., and Mehler, J. (1999). Epenthetic vowels in Japanese: a perceptual illusion. *J. Exp. Psychol. Hum. Percept. Perform.* 25, 1568. doi: 10.1037/0096-1523.25.6.1568
- Fuchs, R. (2016). *Speech Rhythm in Varieties of English: Evidence from Educated Indian English and British English*. Singapore: Springer.
- Fujimoto, M. (2015). “4 Vowel Devoicing,” in *Handbook of Japanese Phonetics and Phonology*, ed. H. Kubozono (Berlin: De Gruyter Mouton), 167–214. doi: 10.1515/9781614511984.167
- Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *J. Exp. Psychol. Hum. Percept. Perform.* 6, 110–125. doi: 10.1037/0096-1523.6.1.110
- Gick, B., and Derrick, D. (2009). Aero-tactile integration in speech perception. *Nature*. 462, 502–504. doi: 10.1038/nature08572
- Gordon, R. L., Schön, D., Magne, C., Astésano, C., and Besson, M. (2010). Words and melody are intertwined in perception of sung words: EEG and behavioral evidence. *PLoS ONE* 5, e9889. doi: 10.1371/journal.pone.0009889
- Guion, S. G., Flege, J. E., Akahane-Yamada, R., and Pruitt, J. C. (2000). An investigation of current models of second language speech perception: the case of Japanese adults’ perception of English consonants. *J. Acoust. Soc. Am.* 107, 2711–2724. doi: 10.1121/1.428657
- Hagiwara, R. (1997). Dialect variation and formant frequency: the American English vowels revisited. *J. Acoust. Soc. Am.* 102, 655–658. doi: 10.1121/1.419712
- Hayashi, M., and Kakehi, K. (1990). “An experimental study on basic perceptual units of speech based on reaction time,” in *Paper Presented at the Spring Meeting of the Acoustic Society of Japan*.
- Hoffer, B. L. (2002). “The impact of English on the Japanese language,” in *Exploring Japanese: On Japanese Enactments of Culture and Consciousness*, ed. R. T. Donahue (Westport: Greenwood Publishing), 263–73.
- Hogg, R. M. (1992). “Phonology and morphology,” in *The Cambridge History of the English Language*, ed. R. M. Hogg (Cambridge: Cambridge University Press), 67–167.
- Inoue, H. (1996). *Nihongo Nikki, Vol. II (Diary of Japanese, Vol. II)*. Tokyo: Bungei Shunjuu.
- Ito, J., and Mester, A. (2015). “7 Sino-Japanese phonology,” in *Handbook of Japanese Phonetics and Phonology*, ed. H. Kubozono (Berlin: De Gruyter Mouton), 289–312.
- Johnson, K., Strand, E. A., and D’Imperio, M. (1999). Auditory-visual integration of talker gender in vowel perception. *JPhon* 27, 359–384. doi: 10.1006/jpho.1999.0100

- Jungers, M. K., Palmer, C., and Speer, S. R. (2002). Time after time: the coordinating influence of tempo in music and speech. *Cog. Proc.* 1, 21–35.
- Kaneko, E. (2006). Vowel selection in Japanese loanwords from English. *LSO Working Papers in Linguistics* 6, 49–62.
- Kentner, G. (2015). “Rhythmic segmentation in auditory illusions-Evidence from cross-linguistic mondegreens,” in *Proceedings of 18th ICPHS*. Glasgow.
- Kubozono, H. (2002). “Mora and Syllable,” in *The Handbook of Japanese Linguistics*, ed. N. Tsujimura (Oxford: Blackwell), 31–61.
- Kubozono, H. (2005). [ai]-[au] asymmetry in English and Japanese. *J. Eng. Linguist.* 22, 1–22. doi: 10.9793/elsj1984.22.1
- Kubozono, H. (2015a). “Introduction to Japanese Phonetics and Phonology,” in *Handbook of Japanese Phonetics and Phonology*, ed. H. Kubozono (Berlin: Walter de Gruyter), 54–93.
- Kubozono, H. (2015b). *Handbook of Japanese Phonetics and Phonology*. Berlin: Walter de Gruyter.
- Labrone, L. (2012). *The Phonology of Japanese*. Oxford: Oxford University Press.
- Ladefoged, P. (2005). *Vowels and Consonants: An Introduction to the Sounds of Languages*. Malden, MA: Blackwell Pub.
- Ladefoged, P., and Broadbent, D. E. (1957). Information conveyed by vowels. *J. Acoust. Soc. Am.* 29, 98–104. doi: 10.1121/1.1908694
- Ladefoged, P., and Johnson, K. (2014). *A Course in Phonetics*. Wadsworth: Cengage Learning.
- LaFeber, W. (1997). *The Clash - U.S.-Japanese Relations throughout History*. New York, NY: Norton.
- Lecumberri, M. L. G., Cooke, M., and Cutler, A. (2010). Non-native speech perception in adverse conditions: a review. *Speech Commun.* 52, 864–886. doi: 10.1016/j.specom.2010.08.014
- Lisker, L. (1999). Perceiving final voiceless stops without release: effects of preceding monophthongs versus nonmonophthongs. *Phonetica* 56, 44–55. doi: 10.1159/000028440
- Loveday, L. J. (1996). *Language Contact in Japan - a Socio-Linguistic History (Oxford Studies in Language Contact)*. Oxford: Oxford University Press.
- Mann, V. A., and Repp, B. H. (1980). Influence of vocalic context on perception of the [j]-[s] distinction. *Percept. Psychophys.* 28, 213–228. doi: 10.3758/BF03204377
- Matsuda, Y. (1985). Cross-over languages: Japanese and English (I). *Kwansei Gakuin Univ. Ann. Stud.* 34, 41–69.
- Matsuda, Y. (1986). Cross-over languages: Japanese and English (II). *Kwansei Gakuin Univ. Ann. Stud.* 35, 47–80.
- McGurk, H., and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature* 264, 746–748. doi: 10.1038/264746a0
- Morrow, P. R. (1987). The users and uses of English in Japan. *World Englishes* 6, 49–62. doi: 10.1111/j.1467-971X.1987.tb00176.x
- Nian, O. S., and Jubilado, R. C. (2011). The linguistic integration of English borrowings in modern Japanese. *Polyglossia* 21, 99–107.
- Nishi, K., Strange, W., Akahane-Yamada, R., Kubo, R., and Trent-Brown, S. A. (2008). Acoustic and perceptual similarity of Japanese and American English vowels. *J. Acoust. Soc. Am.* 124, 576–88. doi: 10.1121/1.2931949
- Nozawa, T., and Cheon, S. Y. (2014). The identification of stops in a coda position by native speakers of American English, Korean, and Japanese. *J. Phon. Soc. Japan* 18, 13–27. doi: 10.24467/onseikenkyu.18.1_13
- Ohata, K. (2004). Phonological differences between Japanese and English: several potentially problematic. *Lang. Learn.* 22, 29–41.
- Ōishi, I. (1992). Japanese attitudes towards the English language during the Pacific war. *Seikei Hogaku* 34, 1–19.
- Olson, I. A. (1997). The dreadful death of the bonny earl of murray: clues from the carpenter song collection. *Folk Music J.* 7, 281–310.
- Otake, T. (2007). “Interlingual near homophonic words and phrases in L2 listening: Evidence from misheard song lyrics,” in *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS 2007)*, Saarbrücken: University of Saarland, 777–780.
- Otake, T., Yoneyama, K., Cutler, A., and Van Der Lugt, A. (1996). The representation of Japanese moraic nasals. *J. Acoust. Soc. Am.* 100, 3831–3842. doi: 10.1121/1.417239
- Peperkamp, S. (2005). “A psycholinguistic theory of loanword adaptations,” in *Proceedings of the 30th Annual Meeting of the Berkeley Linguistics Society*, Berkeley, CA.
- Peperkamp, S., Vendelin, I., and Nakamura, K. (2008). On the perceptual origin of loanword adaptations: experimental evidence from Japanese. *Phonology* 25, 129–164. doi: 10.1017/S0952675708001425
- Pike, K. L. (1945). *The Intonation of American English*. Westport, CT: Greenwood Press.
- Pike, K. L. (1947). *Phonemics - A Technique for Reducing Languages to Writing*. Ann Arbor, MI: University of Michigan.
- Pintér, G. (2015). “The emergence of new consonant contrasts,” in *Handbook of Japanese Phonetics and Phonology*, ed. H. Kubozono (Berlin: Walter de Gruyter), 172–214.
- Scherling, J. (2012). *Japanizing English: Anglicisms and Their Impact on Japanese*. Tübingen: Narr Verlag.
- Scherling, J. (2013). Holistic loanword integration and loanword acceptance. A comparative study of anglicisms in German and Japanese. *AAA-Arbeiten aus Anglistik Amerikanistik* 38, 37–52.
- Shoji, S., and Shoji, K. (2014). “Vowel epenthesis and consonant deletion in Japanese loanwords from English,” in *Proceedings of the 2013 Annual Meetings on Phonology*. Amherst, MA: University of Massachusetts, 1.
- Stanlaw, J. (2002). Hire or fire? Taking AD-vantage of innovations in the Japanese syllabary system. *Lang. Sci.* 24, 537–574. doi: 10.1016/S0388-0001(01)00004-3
- Stanlaw, J. (2004). *Japanese English - Language and Culture Contact*. Hong Kong: Hong Kong University Press.
- Starr, R. L., and Shih, S. S. (2017). The syllable as a prosodic unit in Japanese lexical strata: evidence from text-setting. *Glossa* 2, 1–34. doi: 10.5334/gjgl.355
- Strand, E. A. (1999). Uncovering the role of gender stereotypes in speech perception. *J. Lang. Soc. Psychol.* 18, 86–99. doi: 10.1177/0261927X99018001006
- Tang, K. (2015). *Naturalistic speech misperception* (dissertation). University College London, London, United Kingdom.
- Tsuchida, A. (1995). English loans in Japanese: constraints in loanword phonology. *Working Papers Cornell Phonet. Lab.* 10, 145–164.
- Turk, A., and Shattuck-Hufnagel, S. (2013). What is speech rhythm? A commentary on Arvaniti and Rodriquez, Krivokapić, and Goswami and Leong. *Lab. Phonol.* 4, 93–118. doi: 10.1515/lp-2013-0005
- Vance, T. J. (1987). *An Introduction to Japanese Phonology*. Albany, NY: State University of New York Press.
- Vance, T. J. (2008). *The Sounds of Japanese*. Cambridge: Cambridge University Press.
- Venedlin, I., and Peperkamp, S. (2006). The influence of orthography on loanword adaptations. *Lingua* 116, 996–1007. doi: 10.1016/j.lingua.2005.07.005
- Warner, N., and Arai, T. (2001). Japanese mora-timing: a review. *Phonetica* 58, 1–25. doi: 10.1159/000028486
- Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science* 167, 392–393. doi: 10.1126/science.167.3917.392
- Wassink, A. B. (2001). Theme and variation in Jamaican vowels. *Lang. Var. Change* 13, 135–159. doi: 10.1017/S0954394501132023
- Wright, S. (1954). *The Death of Lady Mondegreen*. Harper’s Magazine, 48–51.
- Zec, D. (2007). “The syllable,” in *The Cambridge Handbook of Phonology*, ed. P. de Lacy (Cambridge: Cambridge University Press), 161–94.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Scherling, Kornder and Kelly. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.