Intonational Structure as a Word-boundary Cue in Tokyo Japanese

Natasha Warner, Takashi Otake and Takayuki Arai

Language and Speech 2010; 53; 107
DOI: 10.1177/0023830909351235

Intonational Structure as a Word-boundary Cue in Tokyo Japanese

Natasha Warner[1,2], **Takashi Otake**[3] **Takayuki Arai**[3]

[1] *University of Arizona, Tucson, AZ, U.S.A.*
[2] *Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands*
[3] *E-Listening Laboratory, Tokorozawa, Japan*
[4] *Sophia University, Tokyo, Japan*

Key words

accentual phrase

boundary

intonation

pitch accent

word segmentation

Abstract

While listeners are recognizing words from the connected speech stream, they are also parsing information from the intonational contour. This contour may contain cues to word boundaries, particularly if a language has boundary tones that occur at a large proportion of word onsets. We investigate how useful the pitch rise at the beginning of an accentual phrase (APR) would be as a potential word-boundary cue for Japanese listeners. A corpus study shows that it should allow listeners to locate approximately 40–60% of word onsets, while causing less than 1% false positives. We then present a word-spotting study which shows that Japanese listeners can, indeed, use accentual phrase boundary cues during segmentation. This work shows that the prosodic patterns that have been found in the production of Japanese also impact listeners' processing.

1 Introduction

When listeners hear connected speech, they must recognize individual words within the speech stream to understand it. At the same time as they are hearing a stream of speech segments, the pitch contour is varying more slowly with the unfolding of

Acknowledgements: We would like to thank Sahyang Kim, Jennifer Venditti, Taehong Cho, Anne Cutler, James McQueen, Lynnika Butler, Tom Bever, Mike Hammond, LouAnn Gerken, Yosuke Sato, and two anonymous reviewers for comments and helpful discussion on this work. We would also like to thank the native listener judges for their assistance and Maarten Jansonius for help preparing stimuli. This work was partially supported by a Grant-in-Aid for Scientific Research (#11610566) to the second author from the Japan Society for the Promotion of Science. Any errors or misinterpretations are, of course, our own.

Address for correspondence. Natasha Warner, Department of Linguistics, University of Arizona, Box 210028, Tucson, AZ 85721–0028, U.S.A.; <e-mail: #60;<nwarner@u.arizona.edu>

Language and Speech 0023-8309; Vol 53(1): 107–131; 351235; DOI:10.1177/0023830909351235
http://las.sagepub.com

the intonational structure. Given the systematicity that intonation demonstrates in languages, one might expect to find effects of intonation on listeners' processing of words. We investigate how listeners might use changes in f_0 or other prosodic information as a perceptual cue to help them find word boundaries during the spoken word recognition process, in Japanese or other languages.

Although it seems easy to recognize individual words in connected speech, speakers do not produce pauses or other obvious boundary cues between most words, and many strings of sounds allow for more than one parse. For example, the sentence "however, email is fast" could also be "how every mail is fast." An f_0 rise at the beginning of "email" could help the listener to identify the word boundary and arrive at the correct parse. Turning to a Japanese example, in the string /moo nizyuuneN/[1] "already 20 years" in Figure 1B below, the pitch rise in /nizyuu .../ could help mark the onset of the second word, and decrease activation of the competitor /oni/ "devil."

There has been extensive past research on how listeners use segmental cues during spoken word recognition, and on intonation. However, past work has rarely related these two topics (Kim, 2004; Welby, 2003, 2007). We will begin by reviewing these two major topics. Mechanisms listeners use to segment continuous speech can be divided primarily into two types: (1) use of phonological knowledge (often language-specific) about word boundaries, and (2) general methods of activation and competition. Examples of the former are phonotactic indications of syllable boundaries in Dutch, German, and English (McQueen, 1998; Weber & Cutler, 2006), vowel harmony in Finnish (Suomi, McQueen, & Cutler, 1997), and phonological patterns in Korean (Warner, Kim, Davis, & Cutler, 2005). To give an example of phonotactic cues, in Dutch the sequence /mr/ cannot occur within one syllable, while the sequence /dr/ can only occur as the onset of a syllable (because /d/ cannot be word-final in Dutch). Dutch listeners use this information in speech segmentation, and recognize a word such as *rok* "skirt" more easily in a sequence /fimrok/ (where the beginning of the word aligns with the beginning of the syllable) than a sequence /fidrok/ (where it does not) (McQueen, 1998). Listeners must use their language-specific knowledge of phonological patterns to do this.

Stress in English is a suprasegmental language-specific cue. English listeners expect stressed syllables to be the first syllable of a new word (they hypothesize a word boundary at a stress), because most high frequency English words begin with a stressed syllable (Cutler & Butterfield, 1992; Cutler & Carter, 1987; Cutler, Mehler, Norris, & Segui, 1992; Cutler & Norris, 1988). F_0 changes are likely to be one important cue to stress, so the fact that stress itself serves as a word-boundary cue supports the hypothesis that intonational cues could be used for word segmentation.

Although listeners use language-specific phonology in segmentation, phonological patterns alone are not sufficient to solve the segmentation problem. Phonotactic cues mark only about 37% of English word boundaries, for example (Harrington, Watson, & Cooper, 1989). Current models of spoken word recognition incorporate general

1 /N/ is the mora-nasal phoneme. Pitch accent will be marked with an apostrophe after the accented mora, but only where it is under discussion and in figures. We use phonemic transcriptions, such as /zyu/ for [dʒu].

methods of activation and competition (e.g., SHORTLIST: Norris, 1994; TRACE, McClelland & Elman, 1986; and episodic models: Goldinger & Azuma, 2003). The SHORTLIST model, for example, uses language-specific phonological information to mark syllable boundaries, which are then considered as likely word onsets in the more general process (Cutler, Norris, & McQueen, 1996; McQueen, Norris, & Cutler, 1994). A shift from activation and competition to a probabilistic Bayesian recognition system (Norris & McQueen, 2008) would not change this combination of language-specific and general recognition mechanisms.

Turning to the topic of intonation, we will summarize the aspects of Japanese intonation that are important for the current topic and then review literature on speech processing and intonation. For Tokyo Japanese, Pierrehumbert and Beckman (1988) find an overall rise–fall pattern at the Accentual Phrase level. An Accentual Phrase contains one or a few content words, along with particles. Pierrehumbert and Beckman analyze Japanese as having an utterance-initial boundary low tone (L%), a phrasal high tone (H) usually at the second mora, an accent HL tone at the pitch accent (if any), and a phrase-final boundary low tone (L%). This results in contours as in Figure 1, with an f_0 rise beginning each Accentual Phrase. In Figure 1, *nizyuuneN izyoo-mo* "more than 20 years even" is a single AP, but the word *moo* has a separate AP.

Traditional literature (e.g., Haraguchi, 1977) describes words with first-mora pitch accent or with a heavy first syllable as beginning with an H tone (e.g., *ma'kura* "pillow," *kookoo* "high school": Vance, 1987). However, Pierrehumbert and Beckman (1988) and Poser (1984) both show that even these words begin with an f_0 rise. Thus, the Accentual Phrase Rise (APR) is consistent as a marker of AP onset.

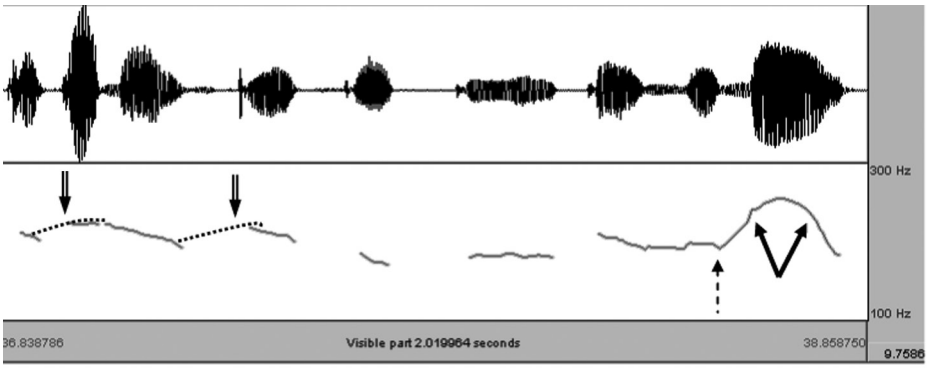
The APR is not lexically distinctive for Tokyo-type dialects (Pierrehumbert & Beckman, 1988; Poser, 1984), so it occurs rather predictably. The basis of the current study is that exactly this predictability of the f_0 rise at the beginnings of phrases should make it an excellent perceptual cue to the location of word boundaries. Looking at Figure 1B, if listeners hypothesize a word boundary at each APR, this would allow them to locate the onsets of *moo* "already," *nizyuuneN* "20 years," and *tsuzukete* "continue," but not *izyoo-mo* "even more." Although APRs mark accentual phrase boundaries rather than word boundaries, they might provide cues to many of the word boundaries. The distribution of tones in Japanese is limited in comparison to English, and there are few rises in the f_0 contour other than APRs. Japanese intonation involves a great many effects that lower f_0 (Pierrehumbert & Beckman, 1988), but only a few that raise it (section 2.21.2 below and Figure 1). This should make APRs very useful boundary cues.

Research on intonation has often focused on production rather than perception or processing. Intonation (consisting for our purposes of the f_0 contour) can be considered as one part of the larger system of prosodic structure, though, and there has been more work on how prosody in general affects listeners' processing than on how the intonational contour itself does. We will focus on the question of intonation affecting processing, with some discussion of broader prosodic effects.

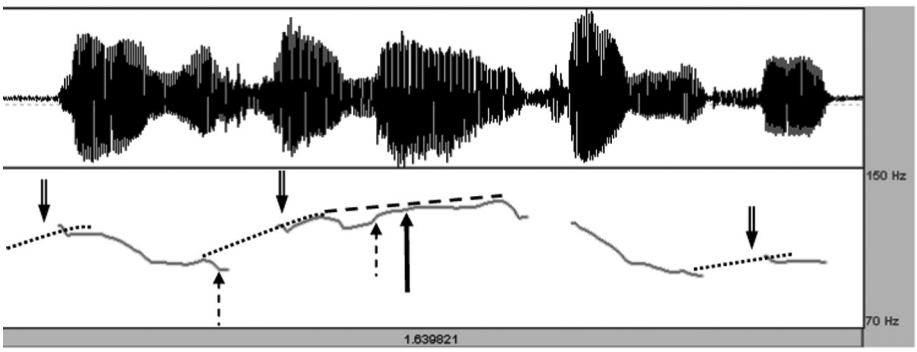
The work that most directly relates intonation to segmentation is that of Welby (2003, 2007) and Kim (2004). Welby (2003, 2007) tests French phrases (e.g., *mécénat* "patronage" vs. *mes sénats* "my senates") that either have or lack an "early rise" LH sequence at a potential word boundary. Listeners more often identify the sequence

Figure 1

Waveforms and f_0 tracks demonstrating both the overall rise-fall pattern for Japanese APs and excluded types of rises. A: The utterance /(a'se-o) kakanai supootu-da-to yuu teN-de-mo/ "There's the fact too that it's a sport where you don't sweat." Dotted curves with double arrows fill in voiceless parts of the signal and trace the pitch rises of the words /kakanai/ and /supootu/. The solid single arrows indicate a rise–fall boundary pitch movement on the final particle /mo/ and the dashed single arrow indicates segmental influence on f_0. B: The utterance /mo'o nizyuuneN i'zyoo-mo tuzu(kete)/ "I've already continued for more than 20 years." Pitch rises in the words /mo'o, nizyuuneN, tuzukete/ are shown by dotted curves and double arrows, while the dashed line with solid single arrow shows the excluded rise from H to HL in /nizyuuneN i'zyoo/ and the dashed single arrows indicate segmental influences. Transcriptions in the figure are phonemic except for audibly deleted segments



A



B

as a single word (*mécénat*) when there is an early rise, suggesting that the contour allows listeners to separate this single word from the preceding context. Examining segmentation at a segmental rather than a syllabic level, Ladd and Schepman (2003) find that the tonal alignment affects English listeners' judgments of which word a consonant belongs to: a later f_0 minimum increases perception of "Ellen Orwell" instead of "Ella Norwell."

Kim (2004) tests whether alignment of tonal contours affects Korean listeners' word segmentation. For example, the real word [mʌɾi] "head" could appear within a longer nonsense string with the possible Accentual Phrase (AP) contour LHLH beginning at or one syllable before [mʌɾi]. Kim (2004) finds easier recognition for words aligned with an AP contour, meaning that they have an AP-final H tone immediately before them. Kim (2004) also finds in a corpus study that the proportion of content words beginning at an AP boundary is extremely high, approximately 90%.

Turning to the effects of prosody in general, Christophe, Peperkamp, Pallier, Block, and Mehler (2004) show that French listeners hearing strings such as *chat grincheux* "grumpy cat" with only a prosodic word boundary between the two words show effects of competition from *chagrin* "sad." However, if a phonological phrase boundary intervenes, it provides such strong cues to the word boundary that there is no competition. Their phonological phrase level is similar to the AP level of Japanese. Gout, Christophe, and Morgan (2004) reach a similar conclusion for infant listeners, and Cho, McQueen, and Cox (2007) similarly show that an Intermediate Phrase (IP) boundary prevents competition for English adult listeners. Thus, either intonation contours or other cues to the crucial boundaries affect segmentation and spoken word recognition.

Salverda and colleagues (Salverda, Dahan, & McQueen, 2003, Salverda et al., 2007) also show effects of prosodic structure on spoken word recognition, but they focus on prosodically-caused segmental differences rather than intonation. There is also substantial work on listeners' use of prosody for syntactic processing (see Cutler, Dahan, & van Donselaar, 1997, and Speer, Warren, & Schafer, 2003, for overviews). Milotte, René, Wales, and Christophe (2008) further show that a prosodic boundary at a similar level to the one tested here is involved in syntactic processing. Finally, work discussed above (Cutler & Butterfield, 1992; Cutler & Carter, 1987; Cutler & Norris, 1988; Cutler et al., 1992; and in addition, Nakatani & Schaffer, 1978) shows that stress in English, however it relates to intonational structure, is a word-boundary cue.

The current study investigates the relationship between pitch rises and word segmentation in Tokyo Japanese. This differs from past work in several crucial respects. Both Korean and French have a wide variety of possible tonal strings within phrases of this size. Kim (2004) finds 13 tone sequences occurring on APs within her Korean corpus (e.g., HH, LL, LHLL, HHLH, LHLH), although some are rare. Welby (2003, 2006) finds six tone contours (e.g. LHLH, LH, LLH) on French APs even within controlled materials. In Japanese, excepting Boundary Pitch Movements (BPMs), the only two possible tone sequences are L% H L% (e.g., /kakanai/ or /supootu/ in Figure 1) and L% H HL L% (e.g., /nizyuuneN i'zyoo-mo/ in Figure 1), as discussed above. Both of these sequences result in an overall rise–(plateau) –fall–(plateau) pattern for the AP, with the differences being in how sharply and when the f_0 falls. Because the possible set of contours is so much more predictable than for Korean or French, intonational cues to AP boundaries have an even greater potential to serve as effective word boundaries.

The flip side of this issue is whether pitch rises that resemble APRs ever begin anywhere other than at word boundaries, potentially leading listeners to posit a word boundary where there is none. A true AP should never begin within a word, so this false-positive problem should be avoided. However, Maekawa and Igarashi (2007) present evidence that particles (which are closely attached to the preceding word)

can sometimes also begin a new AP, and there could be unexpected pitch rises that resemble an APR. The past literature on French (Welby, 2003, 2006, 2007) and Korean (Kim, 2004) does not address this issue of false positives.

In the current study, we use two approaches. First, we present a corpus study on natural speech, to determine whether pitch rises are potentially good word-boundary cues. We examine both correct boundary detection and false positives. Second, we use a word-spotting study to test directly whether listeners can use prosodic cues for Tokyo Japanese word segmentation.

2 Experiment 1: Corpus study

Because much is already known about Tokyo Japanese intonation (e.g., Pierrehumbert & Beckman, 1988; Venditti, Maekawa, & Beckman, in press), it might be tempting to think that we already know how pitch rises will align with word boundaries. However, intonation in spontaneous speech is highly variable compared to what one would expect from carefully controlled experimental materials (Maekawa, Kikuchi, Igarashi, & Venditti, 2002; Venditti, 2005). More importantly, the previous production research on Japanese intonation does not tell us whether intonation can signal word boundaries, because it does not address the issue of spoken word recognition. The past research is about what contours speakers produce, not about whether listeners can use them to help locate words. One can predict based on the production research that many content words should begin at APRs and that APR-like pitch rises should never occur without a word boundary, except for a few special cases (boundary pitch movements (BPMs) and particles that retain accent: Maekawa & Igarashi, 2007). However, the past research does not tell us what *proportion* of content words begin at pitch rises, and thus how useful pitch rises might be during speech segmentation. The use of casual, spontaneous speech is also important: Welby (2003, 2006) used carefully controlled utterances, and Kim (2004) used connected speech partly from a radio corpus, which was on the average more planned than the speech we study.

Unlike Kim's (2004) corpus study, we do not begin by labeling the speech with a tonal transcription system like J_ToBI (Venditti, 2005). Instead, we label *all* pitch rises and classify them as one of several types. This offers more objective evidence of how effective pitch rises could be as a boundary cue, because it does not allow us to use much information other than the f_0 curve. Full tonal labeling makes use of a researcher's knowledge of the language, allowing many factors besides f_0 to influence which rises are considered to be AP boundaries. The current study thus extends the past studies by determining the alignment of objectively identified pitch rises with word boundaries in spontaneous speech, in a third language.

2.1 Methods

2.1.1 The corpus

For this study, we chose a subset of the spontaneous monolog portion of the OGI Japanese corpus (Muthusamy, Cole, & Oshika, 1992). Speakers produced approximately 50 seconds of speech each, recorded over the telephone, after being prompted to tell about themselves. Speakers knew that they were being recorded and that they were not addressing a human interlocutor. All speakers used for this study performed

this task without difficulty, talking comfortably about their hobbies, work, families, experiences in the United States, etc. While this is not a completely natural speech style, most speakers produced quite fluent, casual speech. Most speakers were living in the United States at the time of the recording.

For the current study, we used eight native speakers of Tokyo-type dialects of Japanese. In a previous study (Warner & Arai, 2001), we used 11 speakers constituting a superset of this same corpus. Five native Japanese speakers (including a linguist specializing in Japanese dialects and two phoneticians) judged the speakers, and speakers with any audible influence of English (their L2) on their Japanese, as well as those speaking any non-standard dialect, were excluded for the current study, leaving the eight speakers we used.

2.1.2 Classification of f_0 rises

The recordings were analyzed using the XWaves software (part of XWaves/ESPS from Entropic). Within each recording, we identified all rises in pitch larger than one would expect based on the inherent variability of pitch tracks. F_0 tracks do vary both upward and downward slightly without linguistic relevance, as can be seen in Figure 1. We did not set an arbitrary cut-off (e.g., all increases from one f_0 point to the next greater than 2 Hz must be included, or all rises of a certain size over the span of 50 ms),[2] but instead retained all pitch rises we judged to be large enough to perhaps be meaningful. The category of "possible" pitch rises is discussed below. We then excluded rises or jumps in f_0 of the non-APR rise types in Figure 1 above (final boundary pitch movement, rise from accentual H to accent HL, microprosodic segmental influence), using criteria discussed below.

Other than these, all rises in the f_0 track were retained for analysis. That is, any f_0 rise that could not be ascribed to a very limited set of alternative explanations was treated as if it were an APR, and it was assumed that listeners would hypothesize a word boundary at it if they use APRs as segmentation cues. This was done in order to minimize use of higher-level knowledge of the language, the intonation system, or the particular utterance, to which listeners might not have online access. This decision may slightly exaggerate the number of pitch rises, but will also bias the data against the hypothesis that pitch rises never occur word-medially. During labeling, we encountered only a few pitch rises that this method led us to analyze but which we believed were not truly onsets of APs.

Both continuous rises in f_0 and sudden jumps in f_0 after a voiceless interval were counted as f_0 rises: Japanese words often begin with voiceless obstruents or contain "voiceless" vowels (Vance, 1987), therefore, gaps in the pitch track are to be expected, and they sometimes cover much of the portion of the pitch track that would otherwise be rising (as in each pitch rise of Figure 1).

The criteria for excluding pitch rises were as follows. An f_0 increase was counted as microprosodic segmental influence (exemplified in Figure 1) if it occurred at onset or offset of voicing, or concurrently with a voiced obstruent or the flap /r/. We presume

2 Such criteria often prove rather inaccurate when applied to highly varied spontaneous speech from multiple speakers, at least if one compares them to a human labeler's judgment.

that listeners are quite able to parse microprosodic variations in f_0 as segmental influences, and thus that it is reasonable to exclude these rises from consideration (cf., Welby, 2003, 2006). If f_0 continued to rise slightly immediately after a pitch rise (Figure 1B), and then fell sharply within the following several moras, the rise was excluded as being from a phrasal H to an accent HL tone. This can only occur beginning immediately at the top of a previous pitch rise, with no intervening f_0 fall (it is a slight rise *from* high f_0, not a large rise *to* high f_0), so listeners should have sufficient cues to avoid misperceiving such rises as APRs.

Finally, if a pitch rise occurred at the end of an utterance, we excluded it as a final boundary pitch movement (Figure 1B, and Venditti et al., in press). While research on BPMs suggests that some of them may actually consist of an APR plus other tones, just what tones BPMs are composed of is not at all clear yet (Venditti et al., in press). This category of excluded rise involves the most use of higher-level information (recognizing the end of the utterance), but there are large acoustic differences between most of these pitch rises and a typical APR that might well allow listeners to recognize this type of f_0 rise without reference to the word string. As in Figure 1A, BPMs often involve extreme pitch changes, out of the range of surrounding speech. They often co-occur with extreme lengthening of a single mora for discourse affect (Venditti et al., in press), and the entire BPM (for example a rise–fall) occurs over the span of this single elongated mora (as in Figure 1A). BPMs often occur before pauses, whereas APRs normally occur either after a pause or medially. Thus, although we use knowledge of the utterance to exclude final boundary pitch movements, we believe this distinction is quite robust for listeners.

We sub-classified each remaining rise for whether it followed catathesis or not. In Japanese, after a pitch accent HL tone, the entire pitch range is compressed (catathesis applies), and further APRs after that will be far smaller than those preceding it (Pierrehumbert & Beckman, 1988). Pitch range is reset at the next IP boundary (a higher level than the AP, replaced by Intonation Phrase in Venditti, 2005). If listeners use pitch rises as word-boundary cues, the size of the f_0 increase might be important, and post-catathesis APRs are often rather small, so we wished to investigate these two types of pitch rises separately.

However, in this spontaneous speech data, it is often difficult in practice to determine whether catathesis has applied, or whether perhaps the pitch range has already been reset after it (Venditti, 2005; Venditti et al., in press). Strong focus can create a very large pitch rise even after a preceding pitch accent, while unenthusiastic speech can create small rises even utterance-initially. In Figure 1B, for example, the initial word *moo* "already" follows a pause with a strong boundary, so should not have catathesis applied to it. The following pitch rise, at *nizyuuneN* "20 years," which is the focus of the utterance, is far larger. Therefore, post-catathesis and non-catathesized rises will be combined for some analyses.

A third category consists of very small pitch rises that might be true onsets of small catathesized APs, or might be just variability in the f_0 track. If an increase in f_0 was too small to be sure whether it was linguistically meaningful, it was classified as a "possible APR." The J_ToBI conventions allow a "labeler uncertainty" category symbolized as "2-" specifically for unclear possible APRs (Venditti, 2005), so it is not surprising that we encountered such cases as well.

2.1.3 Locating word boundaries

We labeled all word boundaries in the corpus. Japanese orthography does not place spaces between words, and it can be difficult to determine what counts as a "word." For the current study, we counted the beginning of each morpheme as a word onset except for particles, suffixes, forms of the copula, forms of the auxiliary verb /iru/, and non-initial parts of compounds. (Particles and inflectional material in Japanese follow the content word to which they are bound, and prefixes are very rare.) Example segmentations appear in Table 1. We labeled each word boundary for whether a pause preceded the word or not. This decision was made auditorily, with help from visual inspection of the waveform.

We marked compound-internal boundaries for analysis, as well, in the case of loosely-bound compounds where both parts could be independent words. (These were rare in this spontaneous speech data, but a later project on newscast speech necessitated this category.) Furthermore, we marked filled pauses as such, not as words. In two cases, a string of particles was separated by a substantial pause from the word to which it belonged. These were labeled as post-pausal particles.

Ours is a generous definition of "word," allowing function-word-like material such as *toki* "when" (literally "time") in *sukii-ni iku toki-ni-wa* "when I go skiing" to count as a word. We chose to err on the side of counting somewhat grammatical material as words rather than on the side of a strict definition of "content word" in order to disfavor our hypothesis. By counting the onset of a grammatical morpheme such as *toki* as a word onset, even though it is extremely unlikely to initiate a new AP, we do not exaggerate the potential usefulness of pitch rises as a word-boundary cue. Furthermore, we make minimal assumptions about listeners' ability to separate function words from content words.

2.2 Results

The labeled corpus, pooled across the eight speakers' recordings, included 440 word onsets. (Pauses are common in spontaneous monolog speech, and the words are relatively long because they include suffixes and particles, as demonstrated in Table 1.) There were a further five compound-internal boundaries, six filled-pause onsets, and two post-pause particles. The data constituted 187 non-catathesized pitch rises, 83 catathesized pitch rises, and an additional 38 possible (unclear) pitch rises.

For each word onset, we determined whether it occurred at a non-catathesized pitch rise, a catathesized rise, a possible (unclear) rise, or no rise. Results appear in Figure 2. Overall, 68.0% of all words began at some form of pitch rise. Dividing the data by type of rise, we found that the small possible rises account for only 8.4% of all word onsets. Forty percent of all word onsets are at non-catathesized (quite clear) pitch rises, and a further 19.5% are at catathesized but clear rises. All eight speakers produced reliable rises (catathesized or not) at more than 40% of word onsets.

These results suggest that speakers produce pitch rises at a great many word boundaries, which could make pitch rises very useful as word-boundary cues. However, this includes both utterance-initial and utterance-medial words, and both words after pauses and those medial to the speech stream. Pauses may give an overwhelmingly strong cue to the onset of the next word. Listeners probably do not need any additional cues to a word boundary immediately after a pause other than the pause itself. Pauses

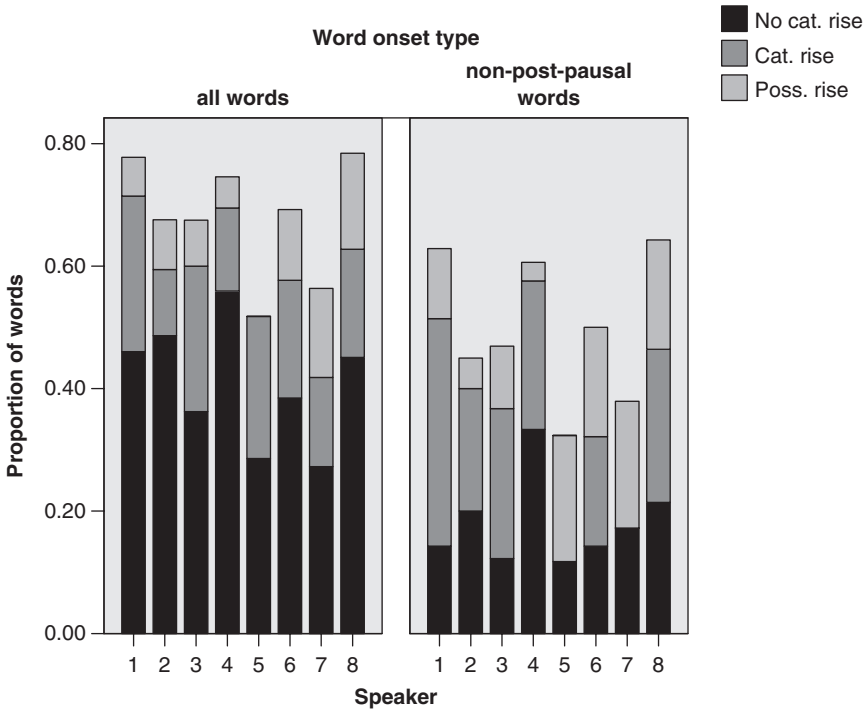
Table 1

Determination of word and other types of boundaries. The morpheme(s) in question appear in bold in the example context, and their nearest English equivalent is also in bold in the translation

Example grammatical morphemes counted as starting new words	*Example context*
toki "when," literally "time"	*sukii-ni iku* ***toki****-ni-wa* **"when** (I) go skiing'
sono 'that" (and other demonstratives)	*sugu-ni yamete simai*, ***sono*** *ato* 'I quit quickly, after **that** ..."
tame "reason"	*ryuugakusee-no* ***tame****-ni* **"because** (I'm) an exchange student"
hoo 'direction"	*minesota-no* ***hoo****-ni kite-iru* "I'm coming to **this place**, Minnesota"
suru "to do" (verbalizer)	*huyu-ni-wa yoku sukii-o* ***site*** "in the winter, (I) **did** skiing a lot"
yuu "to say, to be called"	*watasi-wa itioo ryuugakusee-to* ***yuu*** *koto-de* "For the most part it's **that** I'm an exchange student"
Types/examples of morphemes not counted as starting a new word (any inflected form)	*Example context*
desu, *da* "to be" (copula)	*ima suNde-iru tokoro-wa koNdominiamu-****desu*** **"**The place I'm living in now **is** a condominium."
iru "be doing" (progressive auxiliary)	*ima suNde-****iru*** *tokoro-wa* "The place I'm liv**ing** in now"
-o, *-wa*, *-no*, *-gurai*, *-to*, *-kedo*, *-ne*, *-yo* etc. (particles, including case, sentential, and discourse particles)	*iroNna hito-****wa*** *naNbee-****kara*** *kitari-****toka*** "all sorts of people **(TOPIC)** come **from** South America **and such"**
-no-ni, *-n-zya-nai-desu-keredomo* (combinations of particles and/or inflectional/copular material)	*amerikaziN-to tukiawanaku naru-****N-zya-nai-ka-tte*** "(I) **wonder, doesn't it** get to be **the case that** I don't spend time with Americans"
Examples of morphemes counted as compound-internal boundaries (second portion)	*Example context*
tenisu = *kurabu* "tennis club"	*boku-wa tenisu* = ***kurabu****-ni haitte* "I joined a tennis **club"**
eebee = *buNgaku* "British and American literature"	*gakkoo-de-wa eebee* = ***buNgaku****-o seNkoo site-imasu* "At school, I'm majoring in British and American **literature"**
Examples of filled pauses	*Example context*
eeto, *etoo*, *ee* "uhm"	*tamatama* ***etoo*** *ryuugakusee-ga ooku iru huroa-de* "It's a floor with by chance, **uhm**, lots of exchange students"

Figure 2

Proportion of all words vs. of non-post-pausal words in the corpus occurring with pitch rises, by type of pitch rise and speaker. The types of pitch rises refer to non-catathesized pitch rises, pitch rises after catathesis has reduced the pitch range, and possible (unsure) pitch rises

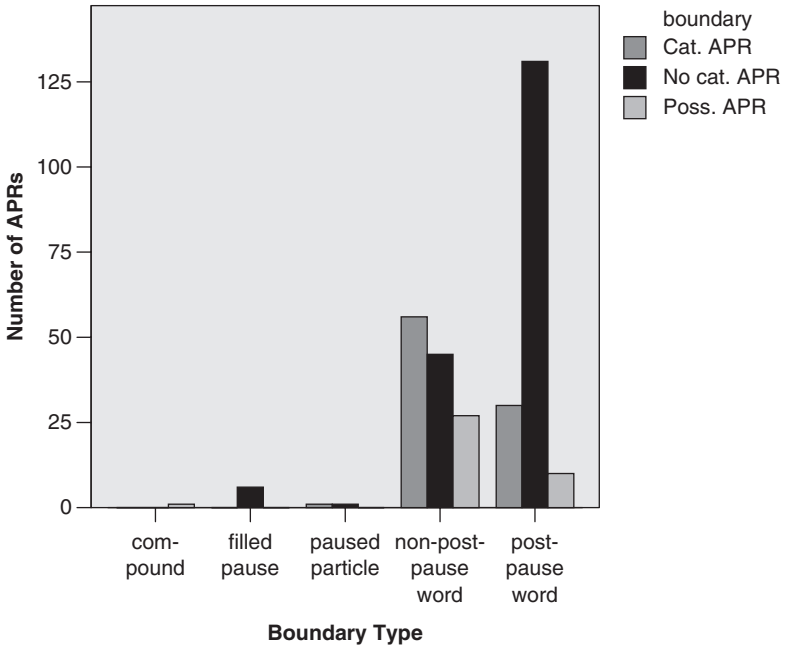


are, of course, also likely to be followed by a new AP. It could therefore be that pitch rises are concentrated at exactly the onsets of the words that are already easy enough to recognize without them, making them of little use for segmentation. We examined the subset of words not following pauses (Figure 2) to rule out the possibility that pitch rises might occur mostly at boundaries where they are not needed as a cue. Even among words that do not follow a pause, 50.0% occur at some type of pitch rise, and 39.5% occur at a clear rise. For all but one speaker, the percentage of non-post-pausal words occurring at a clear rise exceeds 30%.

According to basic assumptions about Japanese intonation (Pierrehumbert & Beckman, 1988), APRs (at least those not involved in BPMs) should only occur at beginnings of words. However, Maekawa and Igarashi (2007) have shown that word-final particles can have a new pitch rise and a pitch accent surprisingly often, and there could be some pitch rises word-medially that are not true APRs. We examined the distribution of included pitch rises relative to all types of boundaries (word, compound-internal, filled pause, etc.) (Figure 3). The overwhelming majority of all types of pitch rise occurs at word boundaries. It is not surprising that the largest quantity of data in any category is non-catathesized pitch rises that occur

Figure 3

Number of pitch rises in the corpus occurring at each type of boundary



at a post-pause word boundary: there is usually a pause before the beginning of an utterance, and an utterance must begin with a non-catathesized APR (Pierrehumbert & Beckman, 1988). Only nine out of the total of 308 total pitch rises (2.9%) occur anywhere other than a word boundary: six occur at a filled-pause onset, two in paused particles, and just one at a compound-internal boundary. Thus, pitch rises at unexpected locations are rare.

2.3 Discussion

2.3.1 Success rate of pitch rises as a word-boundary cue

The data show that a large proportion of words in spontaneous Japanese speech (68% of all words in the corpus) are accompanied by some type of pitch rise. Thus, the rise in f_0 at the onset of an AP could be a strong cue to a word boundary. Even if one examines only words that do not directly follow a pause, assuming that the pause would provide sufficient boundary cues, and even if one excludes small ("possible") rises, the percentage of words co-occurring with a clear pitch rise is 39.5%. This hit rate is similar to that of phonotactic cues in English word-segmentation (37% of words: Harrington et al., 1989), and represents a strong word-boundary cue, which can be used together with general methods of activation and competition.

Kim (2004: 50) finds in her corpus study that an even higher percentage of Korean words, 88.6% of content words, occur at the onset of an AP. There are crucial differences

both in the methods and the languages that affect this comparison, though. Kim's (2004) corpus study examined only content words. In the current study, because of the generous definition of "word" we took, we included rather grammatical morphemes as word onsets, such as *toki* "when" (in Verb-*toki*). This decreases the proportion of word onsets that coincide with pitch rises.

Apart from methodology, AP boundaries may be used for perception in different ways, a topic we will return to in section 4. Furthermore, despite the language-specific differences, it is clear that AP boundaries are very strong potential cues in both Japanese and Korean.

2.3.2 False-positive rate

The results show that if Japanese listeners hypothesize that there is a word boundary at each pitch rise, this would lead to very few false positives. The overwhelming majority of pitch rises (outside of the few exclusion categories) do fall at a word boundary, with only nine anywhere else. This largely confirms the prediction based on theory of Japanese intonation (Pierrehumbert & Beckman, 1988) that pitch rises should only occur at word boundaries. However, it raises the question of what led to the nine anomalous cases.

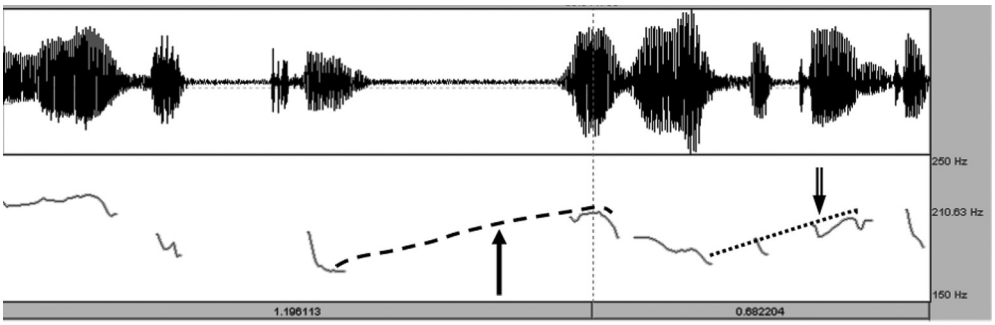
Six of these are pitch rises at filled pauses (e.g., *ee*, *eeto*). These filled pauses would not be labeled as AP boundaries in J_ToBI (Venditti, 2005). They were labeled as pitch rises because they began with higher f_0 than the end of the previous utterance, and our criteria did not allow us to exclude them. However, they are acoustically very different from a true APR, so they are unlikely to create false-positive word boundaries during speech segmentation. They are an artifact of our objective approach to pitch rise measurement.

Of the remaining three anomalous pitch rises, two occur at the beginning of particles that are separated from their preceding word by a pause (Figure 4). These do appear to be true APRs (cf., Maekawa & Igarashi, 2007), and this may parallel historical re-analyses of particles as independent conjunctions. For example, one can begin a colloquial utterance with *kedo* "but," which is usually a particle that follows a verb, rather than an independent conjunction. These particles (e.g., Figure 4) are similarly being treated as a more independent unit. The final case of a pitch rise not at a word boundary is at a compound-internal boundary, in *amerika-buNgaku-no* "American literature (genitive)." In ongoing research we have begun examining newscast speech to gain more insight into compound boundaries.

Overall, it is clear that pitch rises have little potential to cause false-positive word-boundary identifications. Our method inflates the number of false positives by not assuming that listeners have access to a fully parsed intonation structure during processing, but the false-positive rate remains quite low. Kim (2004), in her similar corpus study on Korean, does not discuss false positives. Because she identifies AP boundaries through a complete K-ToBI labeling, her data probably did not include any AP boundaries at non-word boundaries, but the perceptual cues to AP boundaries may indeed lead to false positives that are not identified through K-ToBI labeling. In Japanese, false positives are possible but rare. The results of the corpus study lead to the question of how listeners might use the pitch rise information.

Figure 4

Waveform and f_0 track for /(kaNtorii-)myu'uzikku (pause) -na'do-wa hoto'Ndo/ "country music (pause) and such (I) mostly (don't listen to)." The dashed curve with a single arrow traces the pitch rise at the onset of /-na'do-wa/. The dotted curve with a double arrow indicates the pitch rise on /hoto'Ndo/



3 Experiment 2: Word-spotting

The corpus study in Experiment 1 suggests that pitch rises should be an extremely effective cue for Japanese speech segmentation. In Experiment 2, we utilize the word-spotting method (McQueen, 1996) to determine whether listeners can in fact use this cue (cf., Kim, 2004). We embed a real word such as *kazari* "decoration" after several nonsense syllables, such as *rekereni*. The complete string is not a real word: *rekerenikazari*. The crucial factor, Word Intonation, is the relationship of the word to the prosodic structure. In one condition, this string had a single AP, so that the real word did not begin at an APR ("AP-medial" condition). In the other version, it was produced as two APs, with the second one beginning at the onset of the real word *kazari* ("AP-initial" condition). If listeners hypothesize a word boundary at pitch rises, we predict that they will recognize embedded words such as *kazari* more easily when the word onset coincides with an APR than when it is medial to a single, longer AP.

3.1 Methods

3.1.1 Materials

The target words to be embedded were 24 unaccented three-mora Japanese real-word nouns. Unaccented words are more likely to combine into larger accentual phrases with other content words than words with a pitch accent are (Venditti, 2005), so the use of unaccented words helps to assure that the pronunciations in both conditions are plausible within the Japanese intonation system. Each word was embedded after either three or four moras of phonotactically possible nonsense material, creating items such as *rekerenikazari* (target word *kazari* "decoration") or *sasegiodori* (target word *odori* "dance") (Appendix). The resulting strings did not contain any embedded words, other than the targets, that were longer than a VCV string. Shorter additional embeddings cannot be avoided because of the densely-filled Japanese lexicon.

The target words (but not the nonsense material) were a subset of the items used by McQueen, Otake, and Cutler (2001).

Six additional three-mora real words (unaccented or final accented) were embedded after one or two moras of nonsense material to serve as real-word fillers, and to prevent listeners from listening only for real words beginning at a certain point in the item. Seventy phonotactically possible filler items of four to seven moras containing no embedded words longer than a VCV were also created. There were a further 12 similar practice items, four containing embedded real words.

All items were recorded in a sound-protected booth by the second author (a native speaker of Tokyo Japanese), using a DAT recorder and a high-quality microphone. Recordings were re-digitized at 16,000 Hz. The speaker produced multiple repetitions of each target item both as a single AP and as two separate APs, with the second AP beginning at the embedded real word (Figure 5). Crucially, both of these intonation patterns are quite possible, and in fact common, for Japanese content words, as demonstrated both by the corpus study above and by previous research (Pierrehumbert & Beckman, 1988; Venditti, 2005). Among the fillers without embedded words, approximately one-fourth were realized as a single unaccented accentual phrase, one-fourth as two accentual phrases, and half with more varied pitch accent patterns. All pitch accent patterns for fillers were possible for one- or two-word sequences of Japanese.

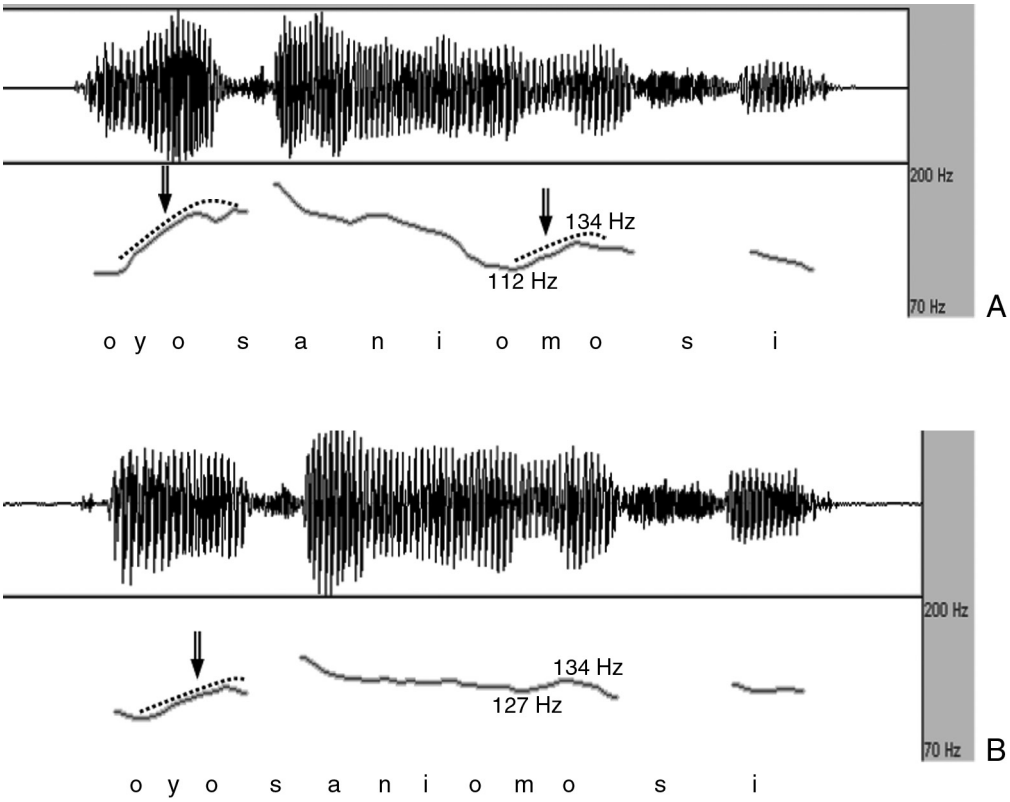
The f_0 contours of all target items were evaluated using XWaves speech analysis software, and the minimum f_0 during the first mora, maximum f_0 during the second mora, and maximum f_0 during the third mora of the embedded word were measured, for both conditions (AP-initial and AP-medial). The third mora was measured because the absolute f_0 peak of an AP is sometimes later than the second mora (Maekawa et al., 2002; Venditti, 2005). Repetitions with larger pitch rises at the embedded word were chosen for the AP-initial condition, and repetitions with clear single-AP contours (gradual fall from the second mora of the item to the end) were chosen for the AP-medial condition. AP-initial embedded words (e.g., Figure 5A) had an average f_0 increase of 17.6 Hz from the first to the second mora, and 17.2 Hz from the first to the third mora. AP-medial embedded words (Figure 5B) had an average f_0 increase of 2.6 Hz from the first to the second mora, and a decrease of 2.2 Hz from the first to the third mora.[3] The effect of Word Intonation on f_0 change was significant, both for first-to-second mora change, $F(1,23) = 182.06$, $p < .001$, and first-to-third mora change, $F(1,23) = 360.11$, $p < .001$, using one-factor within-items ANOVAs.

Although the focus of the current work is on f_0 cues to word boundaries, it is possible that there are other cues to the AP boundary besides f_0. Lengthening at the end of a prosodic unit is well known, with greater lengthening for higher prosodic domains (Byrd, Krivokapić, & Lee, 2006). Kim (2004) finds that lengthening before the boundary is a major cue for AP boundaries in Korean. We therefore took several

3 This does not mean that f_0 was actually increasing from the first to second mora, or decreasing so slightly from the first to third, in these productions. The measurement criterion for both conditions is the lowest f_0 in the first mora and the highest f_0 in the second and third, which will minimize the measured size of any decrease.

Figure 5

Examples of word-spotting stimuli for the word /omosi/ "weight," embedded in /oyosani-omosi/, with f_0 tracks. Dotted curves with double arrows indicate APRs. F_0 measurements are given for the lowest point of the first mora and the highest of the second mora of the target word. A: With two APs, the second beginning at the target word. B: With a single AP and no AP boundary at the target word onset



duration and intensity measures as well,[4] specifically the duration of the last consonant and vowel before the embedded word, of the first consonant (if any) and vowel of the embedded word, and the RMS intensity of the last vowel before and the first vowel of the embedded word (cf., Christophe et al., 2004). We also tested the combined duration of the preceding C and V (the preceding mora) and of the VC spanning the boundary. Criteria for locating boundaries for duration measurements were: offset/onset of F2 of the vowel for vowel–obstruent boundaries; sudden discontinuity in frequencies of energy in the spectrogram for vowel–nasal boundaries; and halfway through the change in F2 frequency for glide–vowel and vowel–vowel boundaries. In a few cases, vowel devoicing/deletion (Vance, 1987) resulted in a [ϕt]-vowel sequence,

4 Using Praat (Boersma, 2001) rather than XWaves.

in which case the onset of [t] was taken to be the offset of frication noise and onset of closure. RMS intensity was the average as calculated by Praat (Boersma, 2001) over the duration identified for each vowel.

One-factor within-items ANOVAs testing the effect of having an AP boundary showed that the only significant non-pitch differences were in duration of the vowel before the embedding, $F(1, 23) = 8.67$, $p < .01$, 86 ms before boundary and 76 ms without boundary, and duration of the VC spanning the boundary, $F(1, 23) = 17.51$, $p < .001$, 112 ms with boundary, 96 ms without.[5] As only eight of the embedded words had an initial consonant, these two measures are closely related.

Stimuli were placed into two lists, with the Word Intonation factor counterbalanced. Thus, each list contained the 24 target items, 12 with the embedded word AP-initial and 12 with it AP-medial. Each list also contained all of the filler items. All items were arranged in a pseudo-random order, with at least one filler containing no embedding immediately before each target-bearing item. The pseudo-random order of the two lists was identical, with only the Word Intonation of the target items changed between lists.

3.1.2 Subjects

Thirty subjects participated in the experiment. All were native speakers of Tokyo-type dialects with no known speech or hearing disorders. Because pitch accent differs widely across dialects in Japanese, only speakers from Tokyo, Kanagawa, Saitama, and Chiba Prefectures were used. Subjects were recruited from linguistics courses at Dokkyo University, which is located in the greater Tokyo area. Subjects received a small amount of course extra credit as a reward for participation. Subjects were divided into two groups of 15 for purposes of counterbalancing the Word Intonation factor.

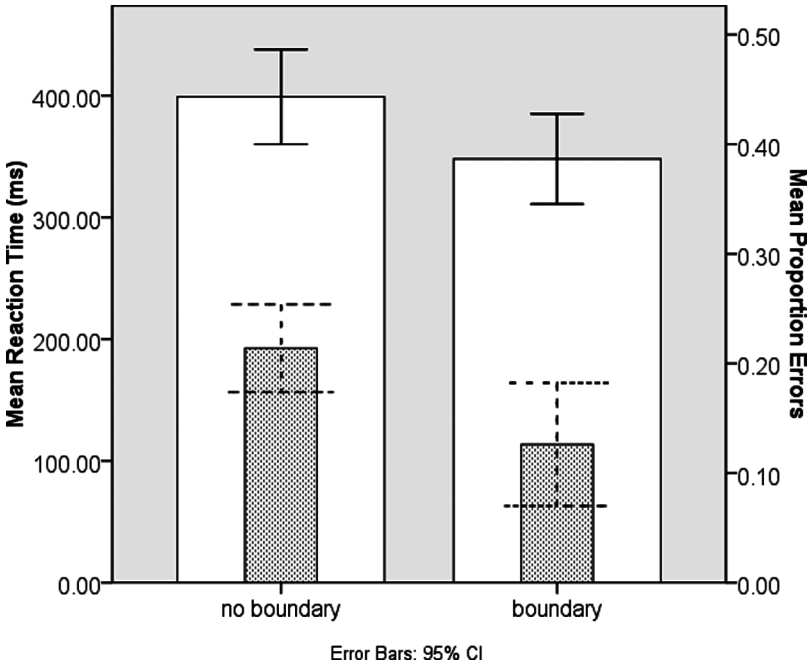
3.1.3 Procedures

The experiment was controlled, and responses collected, by the NESU experiment control software (Wittenburg, Nagengast, & Baumann, 1998). The experiment was presented by the second author, who is a native speaker of Tokyo Japanese, and subjects' only interaction was with him. Subjects were seated in separate sound-attenuating carrels in a quiet room and heard stimuli over headphones. They were tested either individually or in pairs. Subjects were instructed that they would hear a list of nonsense words, and that some of these might contain a real word. Subjects were instructed to push a button on a response box as quickly as possible whenever they heard a real word, and to then say the real word they heard in a low voice into a microphone. Subjects were then presented with the practice list, and then with one of the two experimental lists. Oral responses were recorded onto DAT tape, and the experimenter also monitored oral responses during the experiment. Listeners who were tested in pairs could not hear each others' responses. If subjects did not respond within 2.5 seconds of the onset of a stimulus, the software presented the next stimulus.

5 Results for other measures: preceding mora duration: $F(1, 23) = 3.51$, $p < .08$; preceding C duration: $F < 1$; duration of initial C of embedding: $F(1, 6) = 4.67$, $p < .08$; duration of first V of embedding: $F < 1$; RMS of preceding V: $F(1, 23) = 2.54$, $p > .1$; RMS of first V of embedding: $F < 1$.

Figure 6

RT and error (miss) results for Experiment 2, by boundary condition. White bars show RTs and solid lines show RT confidence intervals. Gray bars show error rates and dashed lines show error rate confidence intervals



3.2 Results

Reaction times (RTs) and error rates (misses) were measured. RTs were adjusted to be measured from the offset of each stimulus.[6] There were no cases of subjects responding, but detecting an incorrect real word. Data from three stimuli (*abata*, *modemu*, *ogura*) were excluded because fewer than 10% of subjects responded to them in at least one condition. (All other stimuli received responses from at least 40% of subjects in both conditions.) RTs faster than 100 ms (approximately 2.7% of the data) or slower than 900 ms (approximately 2.5% of the data) were excluded. No subjects were excluded, as none had extremely high error rates. (The greatest error rate for a subject was 50% in one condition.) Results appear in Figure 6.

By-subjects (*F1*) and by-items (*F2*) ANOVAs were performed on the RTs and error rates (misses), with the repeated measures factor Word Intonation (AP-medial vs. AP-initial). For the by-subjects analysis, a between-subjects control factor of counterbalanced list was also included, but results involving this factor will not

6 Furthermore, the total duration of the embedded words did not differ for the two conditions ($F < 1$).

be reported because they are uninterpretable. Listeners detected embedded words coinciding with an APR (AP-initial words) significantly more quickly and accurately than words medial to an AP (RTs: $F1(1, 28) = 18.52$, $p < .001$, $F2(1, 20) = 6.25$, $p < .03$; Errors: $F1(1, 28) = 8.60$, $p < .01$, $F2(1, 20) = 6.63$, $p < .02$).

3.3 Discussion

These results demonstrate that an AP boundary at the onset of a word makes it easier for listeners to segment that word out of the surrounding speech stream. The AP boundary is realized as a pitch rise along with slight lengthening. This extends Welby's (2003, 2007) and Kim's (2004) findings to a different language with a very different intonational system.

Kim (2004) placed embedded words either at an AP boundary or one syllable after an AP boundary, but unlike the current study, both conditions contained an AP boundary. The latter condition is a possible but unlikely pronunciation for a Korean phrase containing the target word (Kim, personal communication, 2008). The current results demonstrate that listeners can use intonational cues to help them locate word boundaries during online processing even if both pronunciations are quite valid.

The current materials show significant lengthening at the medial AP boundary, which may also provide a perceptual cue. However, the f_0 rise at AP boundaries is quite noticeable, while a duration difference of 16 ms is small relative to most linguistically meaningful duration distinctions. Statistically, the f_0 difference between embedded words with vs. without an AP boundary has a larger effect size (partial eta squared of .88 for preceding mora to first mora f_0 rise) than the duration difference does (partial eta squared of .43 for the VC string, the largest duration effect). The results show an effect of prosody overall rather than specifically the f_0 rise, but the f_0 rise is likely to be the largest perceptual cue to the AP boundary.

4 General discussion

The corpus study in Experiment 1 showed that hypothesizing a word boundary whenever one hears a pitch rise would help listeners detect a large proportion of Tokyo Japanese word boundaries (high hit rate), while only rarely leading them to posit an extra word boundary (low false-positive rate). The word-spotting results in Experiment 2 show that listeners can, indeed, use the cues at an AP boundary. Considering the many, detailed intonational patterns that literature on production of intonation has identified in languages, one would expect listeners to be affected by intonation during online spoken word recognition. The current work demonstrates one way that intonational regularities do affect listeners' processing.

What makes APRs a particularly useful word-boundary marker for Japanese is a combination of several aspects of the Japanese intonational system. First, APRs are not lexically distinctive, nor are they one among many non-distinctive options: all APs must begin with an APR. Furthermore, APRs are largely unaffected by the presence/absence or position of a lexical pitch accent.

This consistency of the APR differs from other languages. Korean (Kim, 2004), French (Welby, 2003, 2006), and English (Beckman & Pierrehumbert, 1986) allow a far

wider range of tonal patterns on APs, so that there is no such consistent marker in the pitch contour for the beginning or end of an AP. Kim (2004) finds that lengthening at the boundary may be a more important cue than f_0 in Korean, and this might reflect the variability of f_0 cues. One would thus expect that it would be harder for Korean listeners than for Tokyo Japanese listeners to locate AP boundaries. Even in Kansai Japanese, words are lexically marked for whether they are low-beginning or high-beginning in f_0 (Pierrehumbert & Beckman, 1988). This means that the f_0 cues to AP boundaries in Kansai Japanese would be more complicated than in Tokyo Japanese.

Another aspect that makes APRs, instantiated as pitch rises, a particularly likely word-boundary cue for Tokyo Japanese is the paucity of other f_0 rises in the intonational system. F_0 can show a rise or rise–fall at the end of an utterance because of boundary pitch movements (BPMs) conveying meanings such as questioning, incredulity, or explanation (Venditti et al., in press). There are very few other causes of f_0 increases in Tokyo Japanese, and none create f_0 contours similar to APRs. This is unlike both English and Korean, where the variety of possible intonational patterns means that potential intonational cues to boundaries will appear at locations other than AP boundaries. Particularly in Korean (Kim, 2004), a likely intonational cue to the AP boundary is a preceding H tone, but many contours contain non-final H tones as well, so there is potential for a large number of false-positive boundary identifications. Kim's finding that lengthening rather than intonation may be the most important AP boundary cue could be related to the inconsistency of tonal cues.

Of course, for Japanese listeners to be able to use APRs as segmentation cues, they would have to be able to map the acoustic pitch rise onto the beginning of the word accurately. APRs normally occur over the first two moras of a word, with the absolute peak in or slightly after the second mora (Pierrehumbert & Beckman, 1988). While the mapping from the rising contour to the word onset is not completely straightforward, native speakers and listeners probably have extensive practice at making it.

Turning now to the size of accentual phrases, the percentage of words coinciding with an AP boundary in the corpus study is a reflection of how many words appear within a single AP in the language. Our results show a large proportion of Tokyo Japanese word onsets coinciding with pitch rises, but fewer than Kim's similar result (2004) for Korean. For another comparison, Kansai Japanese is said to have smaller APs than Tokyo Japanese, with the AP largely coinciding with the word level (Pierrehumbert & Beckman, 1988). Thus, if there are sufficient cues for Kansai listeners to identify AP boundaries, this might be an even more useful word-boundary cue in Kansai than in Tokyo Japanese. Yet, both Korean and Kansai Japanese provide less consistent cues to the location of AP boundaries than Tokyo Japanese, as discussed above. In sum, languages differ both in how easy AP boundaries are for listeners to locate, and in what proportion of word boundaries can be cued by AP boundaries once a listener locates them.

Despite these differences among languages, it is known that listeners use both general segmentation methods and language-specific phonology during spoken word recognition (Cutler et al., 1996; McQueen, 1998; McQueen et al., 1994). The only language-specific word-boundary cue previously suggested for Japanese is the mora: listeners hypothesize that a word could begin at the onset of each mora, but

not in the middle of a mora (Cutler & Otake, 2002; Otake, Hatano, Cutler, Mehler, 1993). That is, on hearing a string such as our stimulus *rekerenikazari*, listeners would hypothesize a new word beginning at each of /r, k, r, n, k, z, r/, but not a word beginning at any of the vowels.

Since Japanese words cannot begin or end in the middle of a mora, mora-based segmentation would in fact allow a listener to locate 100% of all word boundaries (a 100% hit rate). However, this does not represent very good word recognition, as this method would also lead to an extremely high number of false positives: most Japanese words contain several moras, with 4.24 as the average number of moras per word (with particles included in the word) in the corpus in Experiment 1. Mora boundaries might be an excellent word-boundary cue if used in conjunction with pitch rises, which have a reasonable hit rate and a very low false-positive rate. These two language-specific cues combined with general methods of activation and competition could be very effective for Japanese word recognition. In fact, in the discussion above, we have assumed that Japanese listeners locate the beginning of the mora at which an APR begins, rather than the instant in time when the pitch rise begins.

This work in combination with previous research (Kim, 2004; Welby, 2003, 2007; and work on English stress cited above) show that some aspect of intonation can serve as a word-boundary cue in French, Korean, English, and Japanese. One might speculate that intonational patterns can be recruited as word-boundary cues in language in general, as a universal cue. Because intonation patterns vary so widely across languages, and because intonation contours are generally aligned either to stresses and pitch accents or to higher prosodic structures such as APs or IPs, it is not at all clear that pitch changes would align with word boundaries often enough to assist listeners in locating them.

However, many languages' intonation systems have boundary tones associated to the left or right edge of an AP, an IP (Intonational Phrase), or an utterance. Such prosodic units generally begin at word boundaries. Therefore, languages with boundary tones may have a particular intonation pattern at many word boundaries. A cue does not have to mark every word to be useful, so the existence of boundary tones in itself suggests that intonation could become a word-boundary cue.

F_0 rises in particular might perhaps serve as a universal boundary cue despite the language-specific nature of intonation. It is cross-linguistically very common to use a basic rise–fall f_0 pattern for some prosodic unit (Hirst & Di Cristo, 1998; Vaissière 1983a, 1983b), and nearly all the languages in Hirst and Di Cristo's (1998) broad cross-linguistic survey of intonation show this pattern. Pitch rises might therefore provide useful word-boundary cues in many languages, perhaps as a supplement to lengthening or other prosodic effects, although how a pitch rise maps onto a word boundary is very language-specific. Vaissière (1983a, 1983b) mentions that pitch rises are frequently associated with beginnings and falls with endings, and suggests that pitch rises are timed to word boundaries in some languages, but to stresses in others. In languages where rises are timed to stresses, listeners would have to infer word boundaries based on what they know about stress position. Furthermore, pitch changes do not happen instantaneously, and they must be mapped to word boundaries despite consonantal influence, including devoicing. However, recent literature on

tonal alignment to segments (e.g., Atterer & Ladd, 2004) suggests that features of the intonation contour do map onto the segmental string in very systematic ways, and this might provide listeners with enough information to apply intonational cues to word segmentation.

The previous studies on French and Korean (Kim, 2004; Welby, 2003, 2007) both argue indirectly for a surprising degree of cross-linguistic generality of pitch rises as a word-boundary cue. Even though neither language has an intonational system that often places f_0 rises at word onsets as Tokyo Japanese does, both find that a pitch rise can be a word-boundary cue. For French, the f_0 rise that Welby finds to be a word-boundary cue is an "early rise," not the "late rise" timed to the end of a word that is typical of French intonation (Welby, 2003, 2007). Even though Korean does not have a consistent f_0 rise timed to AP-onsets, and Kim (2004) finds a greater effect of cues at the end of the preceding AP, she does find an effect of AP-initial pitch rise on word recognition. Thus, even languages in which the intonational structure does not clearly lend itself to pitch rises being a word-boundary cue may use pitch rises this way. However, it is clear that any conclusions about the generality of pitch rises and segmentation will require much additional research.

To conclude, Experiment 1 shows, based on a corpus of spontaneous Japanese speech, that f_0 rises such as those at the beginning of APs should be good word-boundary cues: they would allow listeners to detect a large proportion of word onsets, while leading to very few false-positive word-boundary identifications. Experiment 2, using word-spotting, confirms that Japanese listeners can, in fact, use acoustic cues to the AP boundary. We suggest, based on comparison with other languages, that this could reflect a universal tendency to align f_0 rises in some way with onsets of prosodic units. Language-specific knowledge of how the particular language's intonation system aligns f_0 rises with boundaries may combine with a universal use of f_0 rises as a boundary cue to facilitate speech segmentation.

References

ATTERER, M., & LADD, D. R. (2004). On the phonetics and phonology of "segmental anchoring" of f_0: Evidence from German. *Journal of Phonetics*, **32**, 177–197.

BECKMAN, M., & PIERREHUMBERT, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, **3**, 255–309.

BOERSMA, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, **5**, 341–345.

BYRD, D., KRIVOKAPIĆ, J., & LEE, S. (2006). How far, how long: On the temporal scope of prosodic boundary effects. *Journal of the Acoustical Society of America*, **120**, 1589–1599.

CHO, T., McQUEEN, J. M., & COX, E. A. (2007). Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics*, **35**, 210–243.

CHRISTOPHE, A., PEPERKAMP, S., PALLIER, C., BLOCK, E., & MEHLER, J. (2004). Phonological phrase boundaries constrain lexical access. I. Adult data. *Journal of Memory and Language*, **51**, 523–547.

CUTLER, A., & BUTTERFIELD, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, **31**, 218–236.

CUTLER, A., & CARTER, D. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language*, **2**, 133–142.

CUTLER, A., DAHAN, D., & van DONSELAAR, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, **40**, 141–201.
CUTLER, A., MEHLER, J., NORRIS, D., & SEGUI, J. (1992). The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology*, **24**, 381–410.
CUTLER, A., & NORRIS, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, **14**, 113–121.
CUTLER, A., NORRIS, D., & McQUEEN, J. M. (1996). Lexical access in continuous speech: Language-specific realisations of a universal model. In T. Otake & A. Cutler (Eds.), *Phonological Structure and Language Processing* (pp.227–242). Berlin: Mouton de Gruyter.
CUTLER, A., & OTAKE, T. (2002). Rhythmic categories in spoken-word recognition. *Journal of Memory and Language*, **46**, 296–322.
GOLDINGER, S. D., & AZUMA, T. (2003). Puzzle-solving science: The quixotic quest for units in speech perception. *Journal of Phonetics*, **31**, 305–320.
GOUT, A., CHRISTOPHE, A., & MORGAN, J. L. (2004). Phonological phrase boundaries constrain lexical access. II. Infant data. *Journal of Memory and Language*, **51**, 548–567.
HARAGUCHI, S. (1977). *The tone pattern of Japanese: An autosegmental theory of tonology*. Tokyo: Kaitakusha.
HARRINGTON, J., WATSON, G., & COOPER, M. (1989). Word boundary detection in broad class and phoneme strings. *Computer Speech and Language*, **3**, 367–382.
HIRST, D., & DI CRISTO, A. (1998). A survey of intonation systems. In D. Hirst & A. Di Cristo (Eds.), *Intonation systems* (pp.1–44). Cambridge: Cambridge University Press.
KIM, S. (2004). The role of prosodic phrasing in Korean word segmentation. unpublished doctoral dissertation, Department of Linguistics, UCLA, U.S.A.
LADD, D. R., & SCHEPMAN, A. (2003). "Sagging transitions" between high pitch accents in English: Experimental evidence. *Journal of Phonetics*, **31**, 81–112.
MAEKAWA, K., & IGARASHI, Y. (2007). Prosodic phrasing of bimoraic accented particles in spontaneous Japanese. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS), Saarbrücken 2007*. Retrieved July 11, 2009, from http://www.icphs2007.de/
MAEKAWA, K., KIKUCHI, H., IGARASHI, Y., & VENDITTI, J. J. (2002). X-JToBI: An extended J-ToBI for spontaneous speech. In *Proceedings of the 2002 International Conference on Spoken Language Processing (ICSLP)* (pp.1545–1548). Retrieved July 11, 2009, from http://www.isca-speech.org/archive/icslp_2002/i02_1545.html
McCLELLAND, J., & ELMAN, J. (1986). The TRACE model of speech perception. *Cognitive Psychology*, **18**, 1–86.
McQUEEN, J. M. (1996). Word spotting. *Language and Cognitive Processes*, **11**, 695–699.
McQUEEN, J. M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language*, **39**, 21–46.
McQUEEN, J. M., NORRIS, D., & CUTLER, A. (1994). Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **20**, 621–638.
McQUEEN, J. M., OTAKE, T., & CUTLER, A. (2001). Rhythmic cues and possible-word constraints in Japanese speech segmentation. *Journal of Memory and Language*, **45**, 103–132.
MILOTTE, S., RENÉ, A., WALES, R., & CHRISTOPHE, A. (2008). Phonological phrase boundaries constrain the online syntactic analysis of spoken sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **34**, 874–885.
MUTHUSAMY, Y. K., COLE, R. A., & OSHIKA, B. T. (1992). The OGI multi-language telephone speech corpus. In *Proceedings of the 3rd International Conference on Spoken Language Processing (ICSLP)* (pp.895–898). Retrieved July 11, 2009, from http://www.isca-speech.org/archive/icslp_1992/i92_0895.html
NAKATANI, L. H., & SCHAFFER, J. A. (1978). Hearing "words" without words: Prosodic cues for word perception. *Journal of the Acoustical Society of America*, **63**, 234–245.

NORRIS, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, **52**, 189–234.

NORRIS, D., & McQUEEN, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, **115**, 357–395.

OTAKE, T., HATANO, G., CUTLER, A., MEHLER, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*, **32**, 258–278.

PIERREHUMBERT, J., & BECKMAN, M. (1988). *Japanese tone structure*. Cambridge: MIT Press.

POSER, W. (1984). *The phonetics and phonology of tone and intonation in Japanese*. Unpublished doctoral dissertation, MIT, U.S.A.

SALVERDA, A. P., DAHAN, D., McQUEEN, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, **90**, 51–89.

SALVERDA, A. P., DAHAN, D., TANENHAUS, M. K., CROSSWHITE, K., MASHAROV, M., & McDONOUGH, J. (2007). Effects of prosodically modulated sub-phonetic variation on lexical competition. *Cognition*, **105**, 466–476.

SPEER, S. R., WARREN, P., & SCHAFER, A. (2003). Intonation and sentence processing. In *Proceedings of the Fifteenth International Congress of Phonetic Sciences*, *Barcelona*, *3–9 August 2003*. Retrieved July 11, 2009, from http://www2.hawaii.edu/~aschafer/SWS_ICPhS03.pdf

SUOMI, K., McQUEEN, J. M., & CUTLER, A. (1997). Vowel harmony and speech segmentation in Finnish. *Journal of Memory and Language*, **36**, 422–444.

VAISSIÈRE, J. (1983a). The search for language-independent prosodic features. In *La percezione del linguaggio* (pp.311–388). Florence: Presso l'Accad. della Crusca.

VAISSIÈRE, J. (1983b). Language-independent prosodic features. In A. Cutler & D. R. Ladd (Eds.), *Prosody: Models and measurements*, (pp.53–66). Berlin: Springer Verlag.

VANCE, T. J. (1987). *An introduction to Japanese phonology*. Albany State University of New York Press.

VENDITTI, J. J. (2005). The J_ToBI model of Japanese intonation. In S.-A. Jun (Ed.), *Prosodic Typology: The Phonology and Intonation of Phrasing*, (pp.172–200). Oxford: Oxford University Press.

VENDITTI, J. J., MAEKAWA, K., & BECKMAN, M. E. (in press). Prominence marking in the Japanese intonation system. In S. Miyagawa & M. Saito (Eds.), *Handbook of Japanese linguistics*. Oxford: Oxford University Press.

WARNER, N., & ARAI, T. (2001). The role of the mora in the timing of spontaneous Japanese speech. *Journal of the Acoustical Society of America*, **109**, 1144–1156.

WARNER, N., KIM, J., DAVIS, C., & CUTLER, A. (2005). Use of complex phonological patterns in processing: Evidence from Korean. *Journal of Linguistics*, **41**, 353–387.

WEBER, A., & CUTLER, A. (2006). First-language phonotactics in second-language listening. *Journal of the Acoustical Society of America*, **119**, 597–607.

WELBY, P. S. (2003). *The slaying of Lady Mondegreen, being a study of French tonal association and alignment and their role in speech segmentation*. Unpublished doctoral dissertation, Department of Linguistics, Ohio State University, U.S.A.

WELBY, P. S. (2006). French intonational structure: Evidence from tonal alignment. *Journal of Phonetics*, **34**, 343–371.

WELBY, P. S. (2007). The role of early fundamental frequency rises and elbows in French word segmentation. *Speech Communication*, **49**, 28–48.

WITTENBURG, P., NAGENGAST, J., & BAUMANN, H. (1998). NESU: The Nijmegen experiment setup. In A. Trapp, N. Hammond, & C. Manning (Eds.), *CIP98 conference proceedings* (pp.92–93). York: CTI Centre for Psychology.

Appendix: Target-bearing items for Experiment 2

All target words are grammatically nouns in Japanese, and all are lexically unaccented.

Target word	*Translation*	*Item*
abata	pock marks	royofuteabata
agura	sit cross-legged	sakoyugiagura
akebi	a type of fruit	nisanasoakebi
atari	hit	mekoyotoatari
ibitsu	crooked	kosananoibitsu
odori	dance	sasegiodori
ogori	conceit	mefunikeogori
ogura	sweet beanpaste	tebemiogura
onaji	same	gofunionaji
owari	end	kiyuchiowari
unaji	nape of neck	isadaunaji
uroko	fish scales	afuteuroko
uwabe	surface	nuyonauwabe
uwasa	rumor	ekowauwasa
garasu	glass	akonigarasu
medaru	medal	rakoyonemedaru
mogura	mole	tsukofuyamogura
nobori	inbound (train)	segofunobori
pedaru	pedal	nugakoyupedaru
omoshi	weight	goyosaniomoshi
kazari	decoration	rekerenikazari
yoroi	armor	natekahoyoroi
asobi	play	tsuseneasobi
modemu	modem	hanokemodemu