

Pitch Accent in Japanese: Implementation by the C/D Model

Osamu Fujimura, DSc

Prof. Emeritus, Department of Speech & Hearing Science
The Ohio State University, Columbus OH, USA
osamu.fujimura@gmail.com

Abstract

In Tokyo Japanese, lexical accent is implemented by pitch pattern control, while phrasal stress patterns, along with pitch variation, convey non-lexical information in discourse. The C/D model represents pitch control by the tonal melody and stress control by the skeletal organization of the utterance. Phonetic implementation of pitch contours is exemplified here for different lexical accent patterns, including interphrasal interaction (catathesis). An example of phonetic variability is discussed. The issue of accent-bearing units (syllable vs. mora) is revisited. A principle of nuclear prominence (PNP) is proposed explaining why pitch drops in the middle of an accented long syllable.

1. Pitch accent in Japanese

Lexical accent in Japanese is generally implemented by pitch patterns, which vary greatly from dialect to dialect. While some dialects do not use any prosodic contrast lexically, Tokyo dialect, among many others, does have contrastive lexical patterns, which, as time functions of F₀, have been widely discussed with acoustic data (Pierrehumbert [1980], Poser [1984], Pierrehumbert & Beckman [1988], Kubozono [1995], Kawakami [1995], Sugito [1996]). The phonological and (abstract) phonetic descriptions are typically couched in terms of a high vs. low binary status, using each mora as the accent-bearing unit. Hattori [1961] (and others following his work) used the accent kernel designated on a mora of a word to indicate a characteristic pitch change, in most dialects a sudden pitch drop, from the pertinent mora to the next. The kernel is given once, at most, for an accentual phrase (Fujimura [1966], McCawley [1968], Hayata [1997]). The pitch change due to the phonological accent kernel can also be a sudden rise, rather than the usual drop in F₀, as in the Sizukuisi dialect, according to Uwano (Uwano [2007]). Some southern dialects of Korean exhibit pitch accent systems similar to Japanese dialects, in that an accent kernel is specified for a word (Hayata, [1997]).

In addition to the pitch change due to the accent kernel, there may be lexically specified phrase-initial syllabic pitch pattern contrasts. Kansai dialects, spoken in Kyoto, Osaka, and elsewhere, exhibit this additional prosodic lexical feature (see for discussion based on data in Sugito [1996] with a report on phonetic variability referring to diachronic changes, Kinsui [1999]). In these dialects, according to the mora-based account, either high or low word-initial pitch is specified for lexical items independently from the designation of the kernel.

There are other pitch changes that are not lexically distinctive, such as the rise from the initial mora to the next in Tokyo dialect, which can be described as a general phrasal

characteristic, noting that even words spoken in isolation must be framed in a phrasal environment in order to be uttered. Thus in Tokyo dialect, a phrase typically starts with a mora in low pitch and the pitch goes up for the next mora, unless the first syllable is accented (Fujimura [1966, 1972], also see *infra* for long syllables). Depending on the discourse function of the utterance, a phonetic phrasal specification evokes different step response functions at phrase boundaries. A case of this is commonly observed in most languages in yes-no questions at the end of the utterance. Such step-response functions are selected according to phrase boundary features and their phonetic realization is sensitive to the given discourse situation as well as dialect.

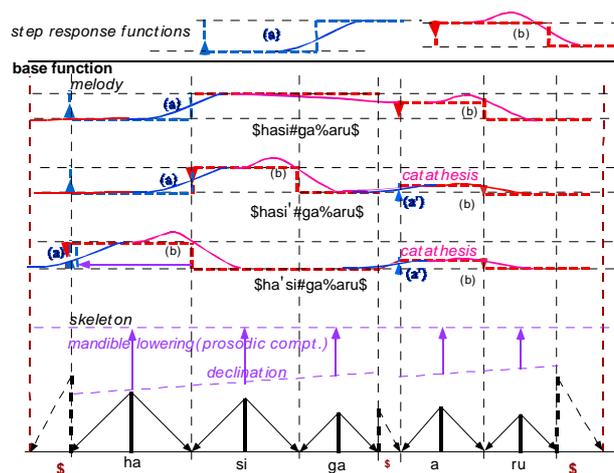


Fig. 1: CD diagram comparing three accent types of the key word: from top, /hasi/ “edge” (no accent kernel), /hasi/ “bridge” (kernel on the second syllable), and /ha’si/ “chopsticks” (kernel on the first syllable), respectively, in ‘Hasi ga aru.’ (There is/are hasi). The last word (copula) ‘aru’ has an inherent accent kernel on the first syllable. In the case of the last two sentences, the pitch dropping effect (b) of this kernel of /aru/ is subject to a substantive reduction (traditionally considered elimination) by catathesis [Poser 1984], as shown by the curve (b’), due to the accent kernel of the word /hasi’/ and /ha’si/, respectively, that is contained in the preceding phrase. Also, the phrase-initial ascending pitch, (a), and its reduced form, (a’) in the case of catathesis, represent the property of (the edge of) the phrase. For the first sentence with /hasi/ with no kernel specified (and therefore no catathesis), no pitch resetting occurs at the small phrase boundary before ‘aru’.

The F₀ time functions, shown here by solid curves, are implementations of the underlying step functions, broken lines, of the phonetic status contours. The bottom panel shows syllable (solid) and boundary (broken) triangles, the size of each representing the syllable or boundary magnitude. In intonation theories (see McCawley,

Pierrehumbert, Poser, among others), the phrase like /aru/ here is called a minor or accent phrase.

A syllable after a boundary (half triangle) reflects the phrase-initial adjustment of the magnitude of the phrase-initial syllable, when the situation evokes catathesis. The syllable magnitude, represented by the pulse height and accordingly the size of the triangle surrounding the pulse, is assumed, in the C/D model, to determine the prosodic component of the jaw opening gesture, as shown in the next-to-bottom panel. These utterances do not involve emphasis but do reflect stress declination as a default phrasal effect (*cf.* Fig. 3 for stress control in English).

The use of step response functions for intonation patterns is similar to Fujisaki's model (Fujisaki & Hirose [1982]), which he more recently used to describe F0 contours of many languages (Fujisaki [2004], Fujisaki *et al.* [2007]). Barbosa's double oscillator model (Barbosa [2006]) may also provide a descriptive framework for such possible F0 patterns, combining two biological oscillating mechanisms, one related to phrasal characteristics and the other syllabic. The response function that we suggest in our model for the accent kernel in Tokyo dialect, based on Poser [1984]'s experimental work, implements a slight rise and subsequent sharp drop of F0 (see Fig. 1, response function (b)). Presumably, such pitch control is implemented by laryngeal muscles (*cf.*, Hirose [1997] for EMG studies, Honda (personal communication) for MRI experimental work).

The pitch fall due to the accent kernel is realized around a crucial time point, which is, *e.g.*, in Tokyo dialect, the end of the (abstract) duration of the kernel-specified syllable if it has no coda (short syllable). If the syllable in question has a coda (*infra*), the crucial time point is roughly in the middle of the syllable duration; the pitch drop occurs around the beginning of the coda manifestation in such a way that the syllable nucleus is perceived not lower in pitch (less prominent) than the rest of the syllable (coda glide). Let us call this constraint the principle of nucleus prominence (PNP). If we assume that the F0 step response is triggered by an excitation pulse at the beginning of the abstract syllable duration (as shown by the red triangle pointed downward in Fig. 1), then we have a common description for long and short syllable types, because the amount of delay of the pitch drop manifestation relative to the beginning of the syllable is fixed according to the step response characteristics, regardless of the total duration of the syllable (for the given tempo specification as a system parameter value). More importantly, apart from phonetic details, this step response description can represent phonetic behaviors in discourse situations and it also suggests a phonologization process that may be responsible for the variability of pitch contour manifestations in different dialects.

At the beginning of a phrase (including a word spoken in isolation) in Tokyo dialect, according to the classical moraic description, the high *vs.* low (HL) phonetic status contour of the phrase begins with a low for the first syllable, as far as the first syllable has no accent kernel specified and also no coda is specified (note that in our system, sokuon is not a syllable coda, as is often assumed, see Fujimura & Williams [1999, 2008]). In other words, if the first syllable is not a long syllable and therefore the phrase-initial mora is low, the pitch switches to high around the boundary between the first and the second morae, according to such descriptions (see the local time function (a) in Fig. 1). This is due to a phonetic implementation rule for a phrase in Tokyo dialect.

A glide, or a combination of glide features, is the only possible coda in a Japanese syllable. The coda glide may be oral (diphthongal palatalization /J/ or monophthongal elongation /H/ of the nucleus) or nasal (/N/, without place specification). If the phrase-initial syllable is a long syllable, *viz.*, when the initial syllable is specified with a glide as its coda, then there is a conflict of the initial rise (see the step response function (a) in Fig. 1) and the kernel implementation (b) as the crucial pitch drop after the nucleus portion of the syllable. The result is, conceptually, a leftward squeeze and reduction of the local time function (a) superimposed onto (b) around the beginning of the first syllable (and of the phrase). If the response function (a) were implemented fully for the long syllable duration according to the mora-based pattern, the coda mora would be more prominent than the nucleus, violating the PNP. Since Japanese speakers (at least in Tokyo) do not use syllable stress enhancement for pitch kernel implementation, syllable prominence is primarily pitch salience as far as lexical characteristics are concerned.

In conformity with the syllable-based phonological description that we assume in general for any dialect, or any language for that matter, the accent-bearing unit is also the syllable, not mora, in our description. The phonetic implementation of the kernel occurs approximately before the coda part of the kernel-bearing syllable under normal situations in Tokyo dialect, in conformity with the PNP above. Corresponding situations in Kansai dialects, in relation to the high *vs.* low starting pitch status, have not been analyzed to date in detail from this point of view. What follows may suggest a possible interpretation of what may appear to be a contradiction to the syllable-based kernel implementation of lexical pitch accent but may not be so if dynamic pitch change is described appropriately.

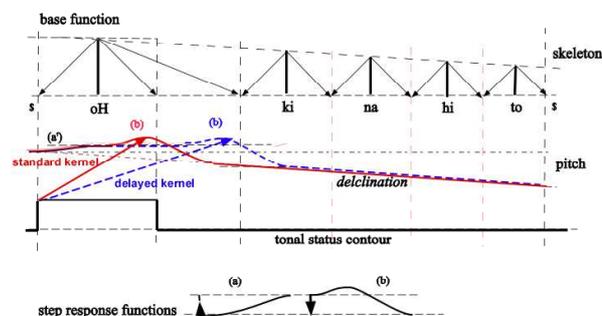


Fig. 2: Phonetic (non-lexical) variability in terms of the timing of the pitch drop effect of the accent kernel. This comparison of two time functions (red solid for standard timing *vs.* blue broken for delayed fall) depicts phonetic variability of the same phrase (Tokyo dialect), and may be interpreted as discrete variability in the moraic high-low representation, HLLL *vs.* HHLL for the same adjective /oH'kina/ (big). In some dialects, a similar pitch curve contrast may be observed as lexically distinctive forms, in terms of the timing of pitch drop, that are interpretable as distinctive moraic patterns.

Details of the F0 contour shape are provided by a phonetic process, which may pertain primarily to delaying a pitch change within a syllable, particularly when a coda is involved (in a long syllable), according to the particular utterance

situation (speaking style, discourse condition, *etc.*). Fig. 2, a CD diagram, depicts a case of stylistic (phonetic) variation of pitch contours, which is commonly observed in Tokyo speakers, perhaps related to a type of emphasis. The phrase, 'ookina hito' (a big person), has a long syllable /oH/ containing the monophthongal vowel elongation /H/ in phrase initial position, and the long syllable in this word carries the lexically specified accent kernel (for explanation of the pseudophonemic transcription, surrounded by slashes, used in this paper, see Fujimura & Erickson [1997]).

Depending on the dialect of Japanese, there can be a lexical phonological contrast between a pitch drop at the middle of a long syllable (end of the nucleus) or at the end of the coda, similar to the phonetic variants in Tokyo dialect as illustrated in Fig. 2. A phonetic system parameter specification scheme can be implemented in the C/D model (see Fujimura [2002, 2008]) to produce two distinct step response functions for the tonal phonetic status contour, according to the speaking style, *etc.* The degree of the delay of pitch fall, as in Fig. 2, may vary continuously according to the phonetic condition, but the delay can be phonologized and become discrete for a particular dialect.

2. Prosody as skeleton and melody

According to the C/D model (Fujimura [2000, 2004, 2007]), the base function of an utterance links many melodies in tonal, vocalic, and other physiological control dimensions to the skeleton, syllable by syllable. Consonantal gestures, as temporally and spatially local perturbations, are superimposed onto the melodic control of the base function in pertinent dimensions.

The skeletal (*viz.*, rhythmic) organization of an utterance is represented abstractly as a train of syllable and boundary pulses. Each pulse is given, phonetically, a scalar magnitude, according to the discourse context, modifying the phonologically (lexically and phrasally) determined discrete (metrical) values. A syllable triangle is computed based on the pulse magnitude for each syllable. Determining the angles of each triangle according to phonological feature specifications of the syllable, the syllable duration, in an abstract sense, is computed by defining it as the base length of the triangle (Fig. 3) (Menezes *et al.* [2003], Bonaventura & Fujimura [2007]). Each boundary pulse also is associated with a triangle, which currently is assumed to have only one side of the vertical pulse (boundary pulse), a half triangle. There are different phonological types of syllables and boundaries, as specified by lexical and phrasal features, and additionally, there is numeric (continuous) modification of the pulse magnitudes according to the utterance context. Unless the phonological type is different (such as long *vs.* short syllables in Japanese, light *vs.* heavy syllable cores and syllables with and without an s-fix in English), all syllables in the same utterance (or a large phrase that can be manipulated independently from other parts of the utterance for contrastive emphasis, *etc.*) are similar triangles (*viz.* with the same angles, see Fig. 3).

Based on lexical prosodic patterns specified and the syntagmatic phonological representation for the linguistic form used for the utterance, phonetic prosodic patterns of Japanese phrases are computed, considering all available utterance information (input specifications for the C/D model), including language/dialect/idiolect, speaker characteristics and idiolectic phonetic disposition, the

speaking style for the particular discourse situation, *etc.* Both skeletal and melodic (including vocalic) patterns of the utterance are affected by these factors. While phonological features use a common descriptive framework, including a more or less universal feature representation scheme (and common feature names), the phonetic implementation process is highly sensitive to the particular language, dialect, speaker idiosyncrasy, *etc.*, just as the speaker's anatomical parameters vary greatly from individual to individual.

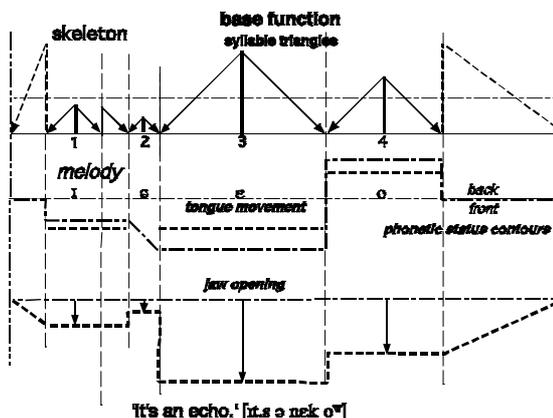


Fig.3: CD diagram for an utterance 'It's an echo.'

Top panel: Syllable triangles showing the size (magnitude) for each syllable. The first syllable contains an s-fix (half triangle). The bottom panel shows a step function for jaw opening (prosodic component): the target values directly reflect syllable magnitudes. The underlying binary status contour for the tongue movement, back/front, is specified as a deviation from the neutral tongue position. This contrast is enhanced phonetically according to the syllable magnitude in excess of a reference value (neutral stress) as shown by the dot-dash line, manifested as a prosodic modulation of vowel quality. The syllable magnitude also controls consonantal gestures, as well as respiratory and other muscular efforts, all linked to the skeletal control of rhythmic patterns.

Stress patterns of an utterance are represented, in the C/D model, by the skeletal structure, which, in turn, is represented by the pattern of syllable-boundary triangles forming a linear series along time. Pitch (F0) patterns are represented by phonetic status contours and laryngeal control variables implement them in the tonal dimensions (Fujimura [2002, 2004]). In English, stress is used for both lexical and phrasal control, and in Japanese, pitch is used for lexical and phrasal specifications. Universally, presumably, both pitch and stress controls are employed at the phonetic phrasal level. Respiratory control of syllable stress (*i.e.*, syllable magnitude) automatically affects F0 through physical mechanisms of phonation, resulting in default F0 variation. Pitch control is provided by laryngeal maneuvers, but default pitch is provided by respiratory and mandibular control due to stress maneuvers. The so-called pitch accent in Japanese lexical prosody involves no stress control (at least in the case of Tokyo dialect) and there is no phonological syllable reduction (but see Uwano [2007], which uses the concept of an inherent syllable strength as a function of phonological features of the syllable for accent rules in certain dialects).

The fundamental frequency (F0) is part of voice quality, and other voice qualities may play crucial roles in prosodic properties of phrasal control.

Japanese lexical prosody is exemplified in Fig. 1, as discussed above, with respect to the manifestation of pitch accent in Tokyo dialect, as interpreted by the C/D model. In this figure, a declination effect, in terms of the syllable magnitude in the phonetic phrasal implementation, is also illustrated. For comparison regarding the roles played by the skeleton *vs.* melodies, English stress control is exemplified in Fig. 3 (American English), involving lexical stress accent and syllable reduction, but no particular emphasis of any word in the discourse is shown in this figure.

Phrase boundaries, in conjunction with discourse properties of utterances such as contrastive emphasis, exhibit very strong phonetic effects, not only on phenomena that are traditionally considered prosodic, but also those considered segmental, in particular vowel quality, *via* mandibular stress control. Phrase initial syllable articulation can be more salient even overriding the effect of phonological syllable reduction, when, for example, contrastive emphasis is employed. Thus, when the word 'America' is emphasized correcting an error in the utterance, 'It's six five seven America Street.', the schwa of the first syllable of the corrected word may have phonetically lower articulation, in terms of the tongue surface height, than the vowel of the second syllable which carries the main stress of the word (see Fujimura [2000]).

References

- Barbosa, P., 2006. *Incurões em Torno do Ritmo da Fala*. Campinas: Pontes Editores.
- Bonaventura, P.; Fujimura, O. 2007. Articulatory movements and phrase boundaries. In M.-J. Solé, P.S. Beddor, M. Ohala (eds.), *Experimental Approaches to Phonology*. Oxford: Oxford University Press, pp. 209-27.
- Fujimura, O. 1966. Kotoba no kagaku (The science of language), the 5th (June 1966 pp. 51-7) of 10 articles in series in the monthly journal *Sizen* (Nature), Tokyo: Chuoo-kooron-sha.
- Fujimura, O. 1972. Kisoron (Fundamental issues). In *Onsei Kagaku* (Speech Science), O. Fujimura (ed.). Tokyo: University of Tokyo Press, pp. 3-91.
- Fujimura, O. 2000. The C/D model and prosodic control of articulatory behavior. *Phonetica* 57, 128-38.
- Fujimura, O. 2002. Temporal organization of speech utterance: A C/D model perspective. *Cadernos de Estudos Lingüísticos, Instituto de Estudos da Linguagem, Campinas* 43, 9-35.
- Fujimura, O. 2004. Prosody: A generalized concept. In G. Fant; H. Fujisaki; J. Cao; Y. Xu (eds.): *From Traditional Phonology to Modern Speech Processing*. Beijing: Foreign Language Teaching and Research Press, pp. 97-110.
- Fujimura, O. 2007. *Onsei Kagaku Genron* (Basic Speech Science). Tokyo: Iwanami Publishers.
- Fujimura, O., Erickson, D., 1997. Acoustic phonetics. In W. J. Hardcastle; J. Laver (eds.), *Handbook of Phonetic Sciences*. Oxford: Blackwell, pp. 65-115.
- Fujimura, O., Williams, J.C. 1999. Syllable concatenators in English, Spanish, and Japanese. In O. Fujimura; B. Joseph; B. Palek (eds.) *Proceedings of LP'98*. Prague: Karolinum Press, pp. 461-98.
- Fujimura, O., Williams, J. C. 2008. Syllables and prosody. *Phonological Studies (Japan)* 14, *in press*.
- Fujisaki, H. 2004. Information, prosody, and modeling – with emphasis on tonal features of speech. In B. Bel, I. Marlien (eds.), *Speech Prosody 2004*, pp. 1-10.
- Fujisaki, H., Hirose, K. 1982. Modeling the dynamic characteristics of voice fundamental frequency with application to analysis and synthesis of intonation. In H. Fujisaki, E. Gårding (eds.): *Preprints of Working Group on Intonation, 13th International Congress of Linguists*, pp. 57-70.
- Fujisaki, H., Gu, W., Ohno, S. 2007. Physiological and physical bases of the command-response model for generating fundamental frequency contours in tone languages. In M.-J. Solé, P. S. Beddor, M. Ohala (eds.): *Experimental Approaches to Phonology*. Oxford: Oxford University Press, pp. 228-45.
- Hattori, S. 1961. Prosodeme, syllable structure and laryngeal phonemes. In *Studies in Descriptive and Applied Linguistics. Tokyo: International Christian University*, Vol. 1, pp. 1-27.
- Hayata, T., 1997. *Ontyoo no Taiporozii*. (Typology of Tonal Systems.) Tokyo: Taisyukan Publishers.
- Hirose, H., 1997. Investigating the physiology of laryngeal structures. In W. J. Hardcastle, J. Laver (eds.): *Handbook of Phonetic Sciences*. Oxford: Blackwell, pp. 116-36.
- Kawakami, S., 1995. *Nihongo Akusento Ronsyuu*. (Collected Papers on Japanese Accent). Tokyo: Kyuuko Shoin Pub.
- Kinsui, S. 1999. Accent on special morae in the Osaka dialect: A study based on data from the CD-ROM Accen Dictionary of Spoken Osaka and Tokyo Japanese. In Onsei Bunpou Study Group (eds.), *Speech and Grammar II*, Tokyo: Kuroshio Publ. pp. 173-99.
- Kubozono, H., 1995. Perceptual evidence for the mora in Japanese. In B. Connell, A. Arvaniti (eds.) *Papers in Laboratory Phonology IV: Phonology and Phonetic Evidence*. Cambridge: Cambridge University Press.
- McCawley, J. D. 1968. *Phonological Component of a Grammar of Japanese*. The Hague: Mouton.
- Menezes, C., Pardo, B., Erickson, D., Fujimura, O. 2003. Changes in syllable magnitude and timing due to repeated correction. *Speech Communication* 40, 71-85.
- Pierrehumbert, J. B. 1980. *Phonology and Phonetics of English Intonation*. MIT Doctoral Dissertation.
- Pierrehumbert, J. B., Beckman, M. E. 1988. *Japanese Tone Structure*. Cambridge, MA: MIT Press.
- Poser, W. 1984. *Phonetics and Phonology of Tone and Intonation in Japanese*. MIT Doctoral Dissertation.
- Sugito, M. 1996. *Osaka-Tokyo Accent Dictionary CD-ROM*. Tokyo: Maruzen Pub.
- Uwano, Z. 2007. Syllable and Mora in the Tokyo Japanese Accent. *Handout at the 2007 Annual Meeting of the English Linguistics Society of Japan, Symposium on Japanese Prosody*.