Working Papers of the Cornell Phonetics Laboratory
No. 1 (December 1983), pp. 1-16

SRS Pitch Rules for Japanese[1]

Mary E. Beckman, Susan R. Hertz, and Osamu Fujimura[2]

Abstract. SRS (Speech Research System) [Hertz 1982] has been used recently to write a set of phoneme-based rules for Japanese. This rule set tests several hypotheses about Japanese phonology. In particular, it tests the usefulness of a hierarchical phrase-structure analysis for generating pitch patterns. The user demarcates the phrases in the romanized input string with boundary symbols, which have associated boundary features. These boundary features, together with user-provided accent marks, are used to generate a segment-by-segment specification of pitch features, such as [low] (low-pitched) and [acc] (accented). Synthesizer parameter rules then refer to the features of the boundaries and of the intervening segments to produce various parts of an utterance's pitch pattern, such as phrase-final lowering or the post-accentual fall. Our SRS pitch rules for Japanese have produced natural-sounding intonational and accentual patterns in a large variety of utterances.

Introduction

Synthesis by rule of prosodic patterns is important not only because it strongly affects the quality of the resulting speech, but also because it offers an experimental means of testing phonological theories. In this paper we will describe our preliminary model for synthesizing pitch

----

[1] A shorter version of this paper was presented at the 105th Meeting of the Acoustical Society of America, May 1983.

[2] Osamu Fujimura, Bell Laboratories, 600 Mountain Ave., Murray Hill, NJ 07974.

patterns in Tokyo Japanese. Japanese is particularly suitable for developing such a model because it allows us to study the effects of phrasal units on pitch patterns unobscured by such complications as the durational effects and large repertoire of pitch contours that can accompany stresses in, for example, English.

First we will describe quite generally our posited phrasal units and their effects. Then we will show how these effects are implemented by the specific rules we developed using SRS, Hertz's language-independent synthesis rule development system [Hertz 1982].

The Accentual Phrase

The smallest phrasal unit that affects the pitch contour is the accentual phrase, which we represent in the traditional manner as patterns of pitch highs and lows determined by the placement of an accent kernel [Miyata 1927, 1928; Hattori 1954].[3] (See Figure 1.) An accentual phrase

----

[3] Our accentual phrase corresponds roughly to the unit that McCawley [1968] calls a minor phrase (Japanese bunsetsu or gosetsu--see Arisaka 1941, Hattori 1947, 1949). However, our treatment of accent suppression (see discussion under "Accent Suppression") often makes our accentual phrase a somewhat larger unit, encompassing two or more minor phrases.

There is an extensive body of literature describing accentual patterns at the level of the minor phrase, including various attempts to predict the accentual pattern of a phrase from the underlying patterns of the component words. The three main issues in this area are the patterns of compound words, the patterns of inflected forms such as verbs and adjectives, and the accentual characteristics of particles [cf. Akinaga 1958, Kawakami 1966, McCawley 1968]. However, these questions lie outside the scope of this paper, since we take as our input the surface forms resulting from any modifications to accent kernel placement made during the word-formation process.

We chose a hybrid analysis involving the high and low levels in addition to the accent kernel, because it is convenient to implement in our present segment-based rule set. Although the high-low analysis forces an oversimplified approximation to such phenomena as accent subordination, it allows for a surprisingly accurate differentiation of perceptually relevant phrasal types.

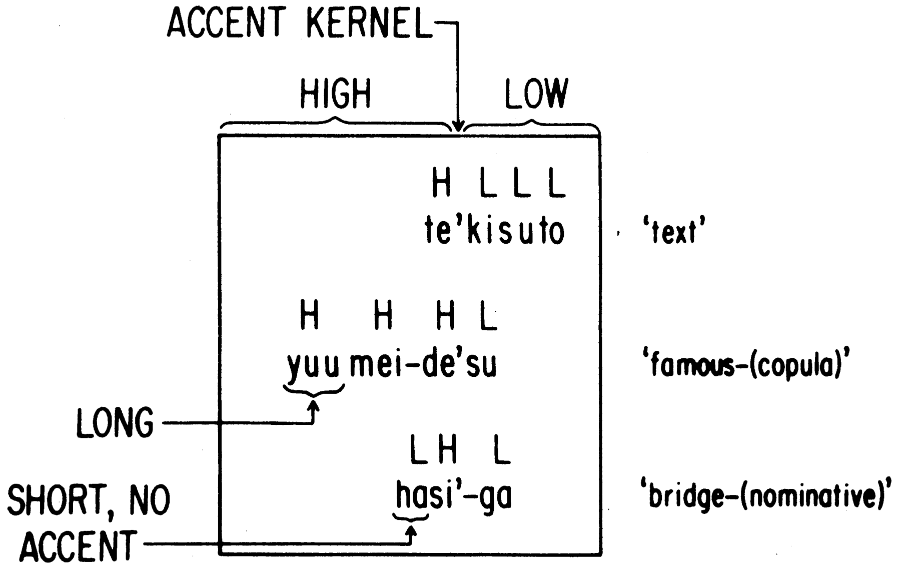

Figure 1. The Accentual Phrase

L H L	hasi'-ga	'bridge-(nominative)'
H L L	ha'si-ga	'chopsticks-(nominative)'
L H H	hasi-ga	'edge-(nominative)'

Figure 2. Accent Contrasts

has at most one pitch drop, located at the accent kernel. Everything that follows this pitch drop is low in pitch, and everything that precedes it is high, with one exception--if the phrase-initial syllable is both short and unnaccented, it is low.[4]

Accent Contrasts

The placement of the accent kernel that determines the highs and lows varies depending on the particular lexical items in the phrase. For example, an accentual phrase consisting of the noun plus nominative particle *hasi-ga* can have one of three contrasting accentual patterns, as illustrated in Figure 2. It can have accent on the second syllable, as in *hasi'-ga*, 'bridge'; it can have accent on the first syllable, as in *ha'si-ga* 'chopsticks'; or, it can have no accent, as in *hasi-ga* 'edge.'

Accent Suppression

Not all lexical accents are realized, however. For example, in most contexts, when a short two-accent sentence, such as *hasi'-ga a'ru*, is imbedded into another sentence as a relative clause, the second pitch fall, in this case the one on *a'ru*, is suppressed (see Figure 3). We deal with accent suppressions of this sort by assuming the lexical item with the suppressed accent to be contained in the same accentual phrase as a preceding item with realized accent. All accent kernels but the first are then ignored.[5] In utterance 2, for example, we would analyze as one accentual phrase the entire sequence *hasi'-ga-a'ru*, causing the accent on *a'ru* to be ignored.

[4] The limitation of this phrase-initial low to short syllables is discussed in Hattori [1954] and Fujimura [1966].

[5] There is some evidence that this suppression is not a complete phonological deletion, but a variable phonetic reduction [e.g., Kawakami 1977]. In addition, there seem to be some special peculiarities when the last element in the accentual phrase is a monosyllabic particle [see Kawakami 1966]. Our model, however, is a first approximation for testing only the gross effects.

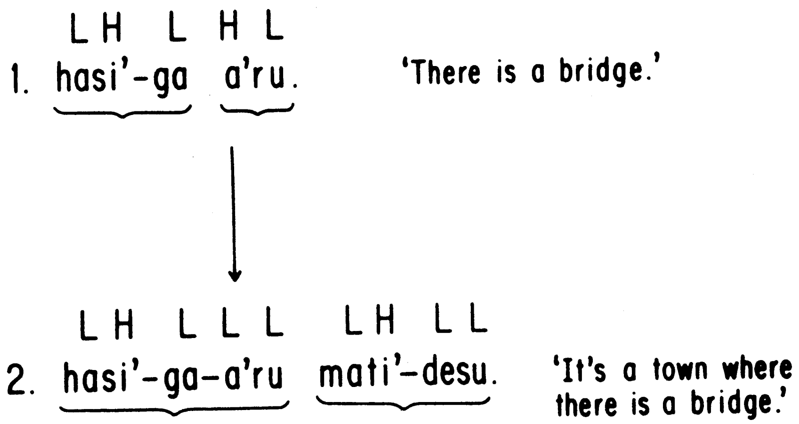
LH L H L
1. hasi'-ga a'ru.
'There is a bridge.'
LH L L L LH LL
2. hasi'-ga-a'ru mati'-desu.
'It's a town where there is a bridge.'

Figure 3. Accent Suppression

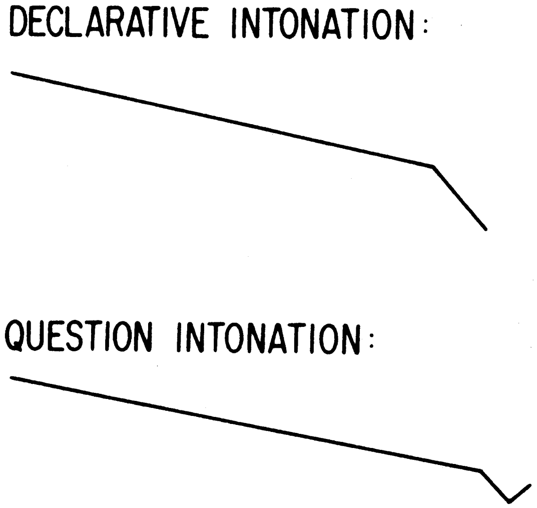

Figure 4. The Sentence

The Sentence

At the other end of the scale from the accentual phrase, we posit--as as the largest phrasal unit in our current scheme--the "sentence."[6] A sentence is the domain of such basic "intonational patterns" as declarative intonation or question intonation (see Figure 4). An intonational pattern has two components. The first is an overall declination; fundamental frequency starts high and gradually falls throughout the span of a sentence. The second is terminal effects, such as the sharp downward curve for a declarative sentence, or the slight upward curve for a question.

The Sentence [with examples]

For simple, short utterances, the two levels "accentual phrase" and "sentence" usually suffice. The highs and lows of the constituent accentual phrases are simply superimposed onto the overall intonational pattern, as illustrated in Figure 5.

The Major Phrase

The pitch contours of complex or longer utterances, however, cannot be generated without assuming an intermediate phrasal unit between the sentence and the accentual phrase.

We currently set up one level of intermediate unit--the "major phrase"--illustrated in Figure 6. The main effect of a major phrase boundary is to reset the basic declination line, starting it again sentence-medially at a fixed value slightly lower than the sentence-onset value. A major phrase, like a sentence, can have various terminal effects, such as a phrase-final rise, fall, or leveling-off. The sentence-final effects discussed above, then, may be considered a special class of such phrase-final adjustments.

----

[6] In future implementations, we may add a unit larger than the sentence--perhaps the "paragraph"--in order to systematically control variations in starting pitch values and other parameters as they relate to relationships among sentences in a discourse.

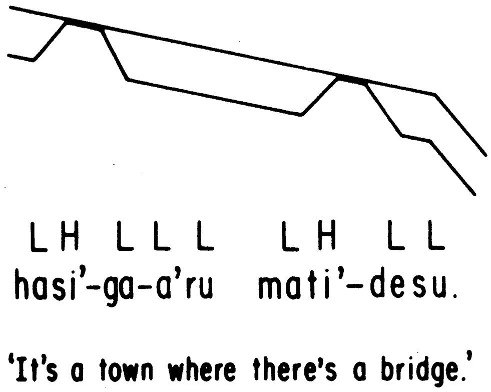
LH LLL LH LL
hasi'-ga-a'ru mati'-desu.
'It's a town where there's a bridge.'

Figure 5. The Sentence

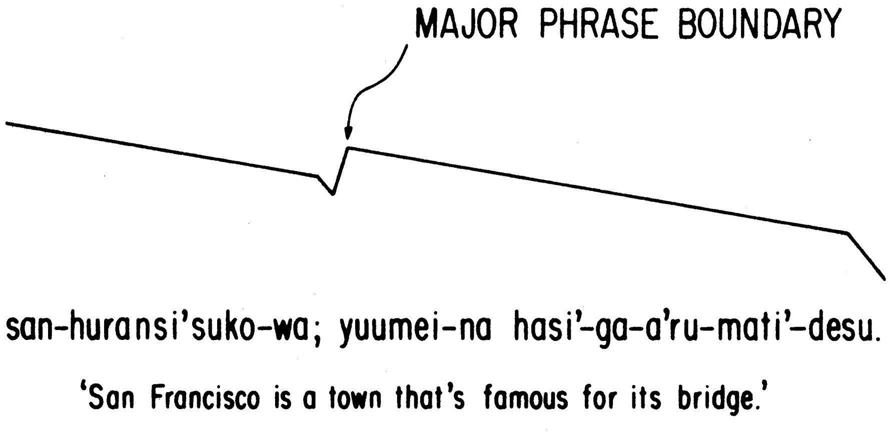
MAJOR PHRASE BOUNDARY
san-huransi'suko-wa; yuumei-na hasi'-ga-a'ru-mati'-desu.
'San Francisco is a town that's famous for its bridge.'

Figure 6. The Major Phrase

Annotated Input Text

The first step in generating the fundamental frequency pattern for an utterance is to annotate a romanized input representation with the appropriate phrasal boundaries and accent marks. In the sample sentence shown in Figure 7, for example, the apostrophes represent the accent kernels, the semicolon represents the end of a major phrase, and the spaces represent the boundaries of accentual phrases. The hyphens mark the boundaries of what might be called "words." These word boundaries affect durational and segmental patterns, but not the FØ contour.

Symbols and Features

The annotated romanized text is then rewritten by a set of "conversion rules" into a string of symbols, which are defined in terms of features to be used by later rules. For example, the semicolon after wa in our sample sentence is rewritten as the symbol [$], which has the associated feature [majp] (for "major phrase boundary"), as illustrated in Figure 8. Similarly, the boundary after na is defined by the feature [accp] (for "accentual phrase boundary"), and the segment [i] of hasi' is defined by the features [nuc] (for "syllable nucleus") and [vh.3] (for "third degree of vowel height"--that is, "high vowel"). Notice how the accent symbol following the [i] adds the prosody feature [acc] (for "accented") to the features inherent in the [i]. Note, too, that at this stage, the [a] of a'ru is also accented.

Accent-Related Features

The output of the conversion rules is then passed to a set of "feature-modification rules." These rules first add the features of lower-level boundaries to those of higher-level boundaries, so that, for example, an accentual phrase boundary would acquire also the features for a word boundary, and similarly, a major phrase boundary would acquire also the features for an accentual phrase boundary. The rules then use these boundary features to assign accent-related features to the utterance segments. For example, Rule 1 in Figure 9 assigns the feature [rise] (for "pitch rise") to a syllable nucleus that is [-long] and [-acc], when it follows an accentual phrase boundary and an optional intervening consonant--that is, when it is phrase-initial. Rule 2 then propagates the feature [low] to all segments in the same accentual phrase to the right of the first accented segment. More specifically, a segment becomes low when it follows

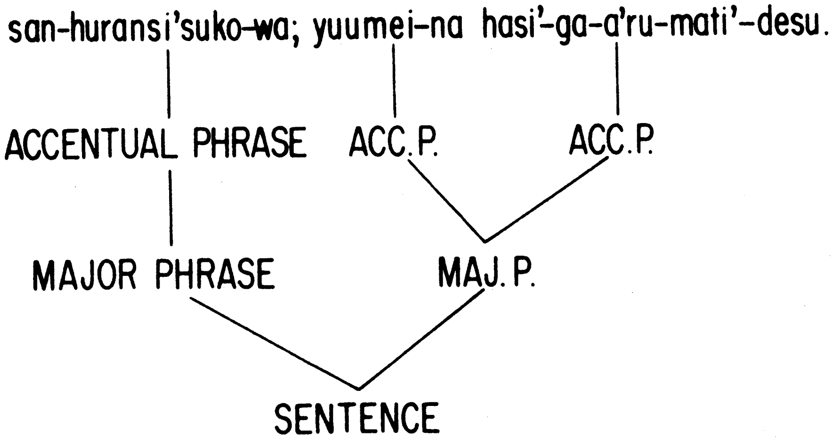

Figure 7. Annotated Input Text

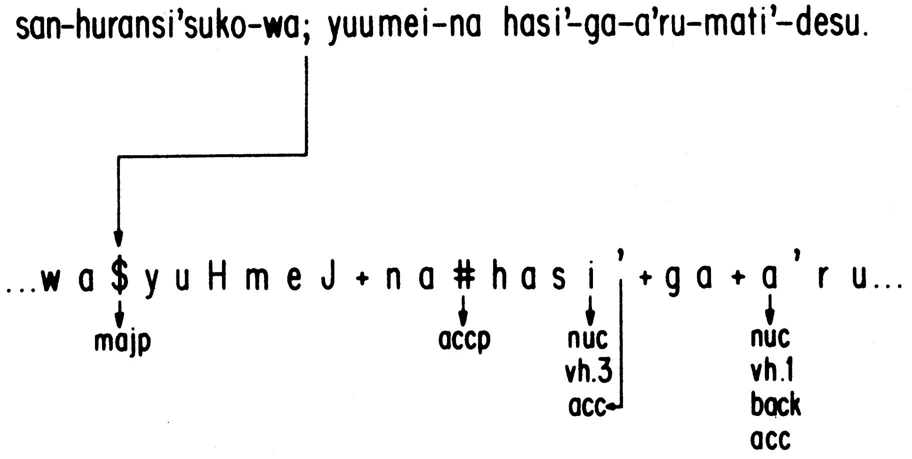
san-huransi'suko-wa; yuumei-na hasi'-ga-a'ru-mati'-desu.
...w a $ y u H m e J + n a # h a s i ' + g a + a ' r u...
majp
accp
nuc
vh.3
acc
nuc
vh.1
back
acc

Figure 8. Symbols and Features

another segment that is either accented or low, so long as no accentual phrase boundary intervenes. Rule 3 uses the feature [low] assigned by Rule 2. It deletes the feature [acc] for any [low] syllable, in this case, suppressing the accentual pitch drop in a'ru.[7]

The Topline

This completed segmental representation is then passed to a set of parameter rules, which generate the appropriate FØ patterns.[8] The rules first produce a declining topline for each major phrase, by assigning a fixed percentage decrement to each segment, as shown in Figure 10.

Terminal Effects

The next rules produce the appropriate terminal contours for major phrases. The rule shown in Figure 11, for example, implements a phrase-final lowering in the last segment of non-sentence-final major phrases. It specifies a sharp linear descent that starts from the topline value at the beginning of the segment--that is, at (.0)--and terminates at the end of the segment--that is, at (.99)--with a value 90% of the topline value at that point.

----

[7] As stated in note 5, this implementation of accent suppression may be an oversimplification. A more accurate rule might merely reduce the accentual effect of a kernel following another kernel in the same accentual phrase, with the degree of reduction a variable function of the distance between the kernels [see Sagisaka and Sato 1983].

[8] The parameter rules described in this paper have been simplified somewhat so that irrelevant detail will not distract from the discussion of our pitch contour implementation. For example, the rule for the phrase-final lowering shown in Figure 11 collapses aspects of several more complicated rules that interact to handle all of the terminal effects. Also, the exact placement of the rise and fall illustrated in Figure 12 varies depending on the segmental composition of the affected syllables, and in some cases there is a slight boost on the accented syllable before the fall.

1. [nuc-long-acc]→[rise]/[accp]([con])_

2. [seg]→[low]/[{acc|low}]~⟨[accp]⟩_

3. [low]→[-acc]

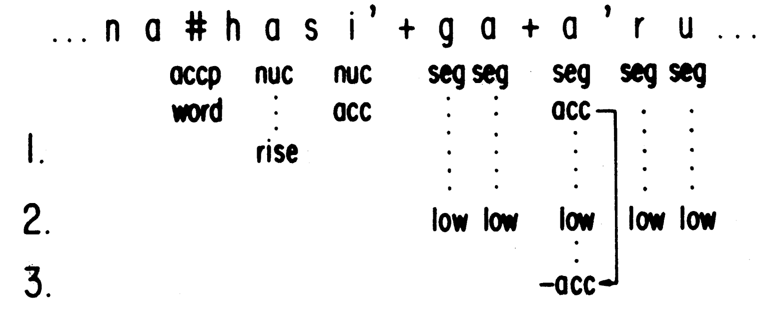

Figure 9. Accent-Related Features

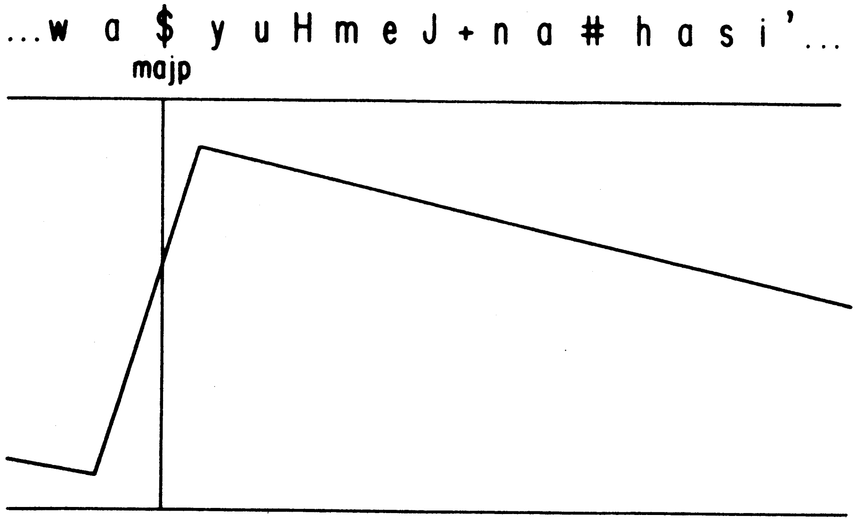

Figure 10. The Topline

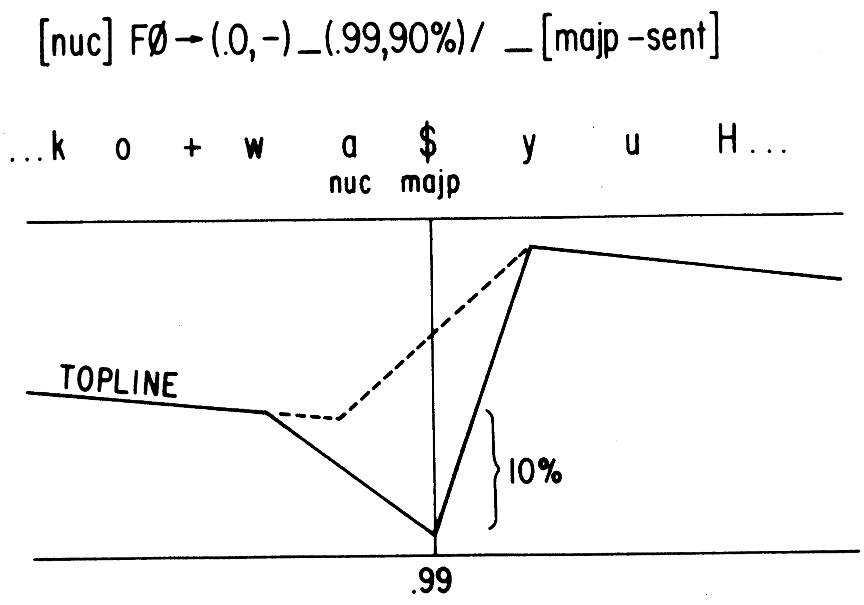

Figure 11. Terminal Effects

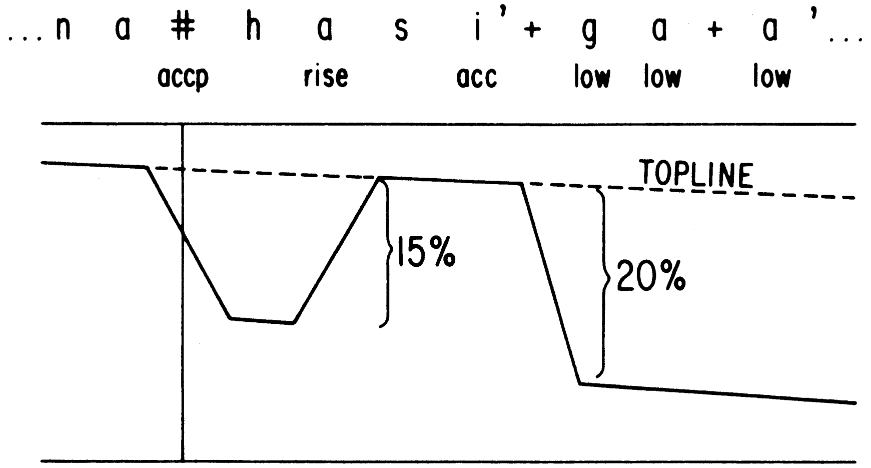

Figure 12. Accentual-Phrase Effects

Accentual-Phrase Effects

Next, the highs and lows of accentual phrase patterns are generated. (see Figure 12). First, the FØ of any syllable marked [rise] is lowered by 15% of the value assigned by the topline rules, yielding the initial rise from low to high for short, unaccented phrase-initial syllables. Then, the segments marked [low] are lowered by 20% of their topline values to produce the accentual pitch drop and subsequent sequence of low-pitched syllables.

The Pitch-Scale Transform

Finally, the values generated by the parameter rules are fed into a function that raises the lower end of the frequency scale, as illustrated in Figure 13. This pitch-scale transform prevents the FØ values from descending to an unnaturally low frequency range. It also has the effect of flattening out the rises and falls at later, lower FØ values.

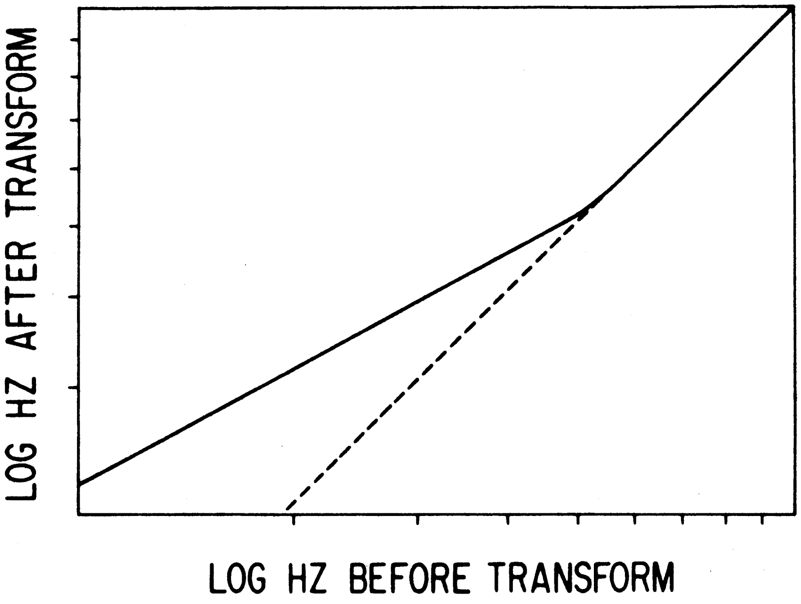

Figure 13. The Pitch-Scale Transform

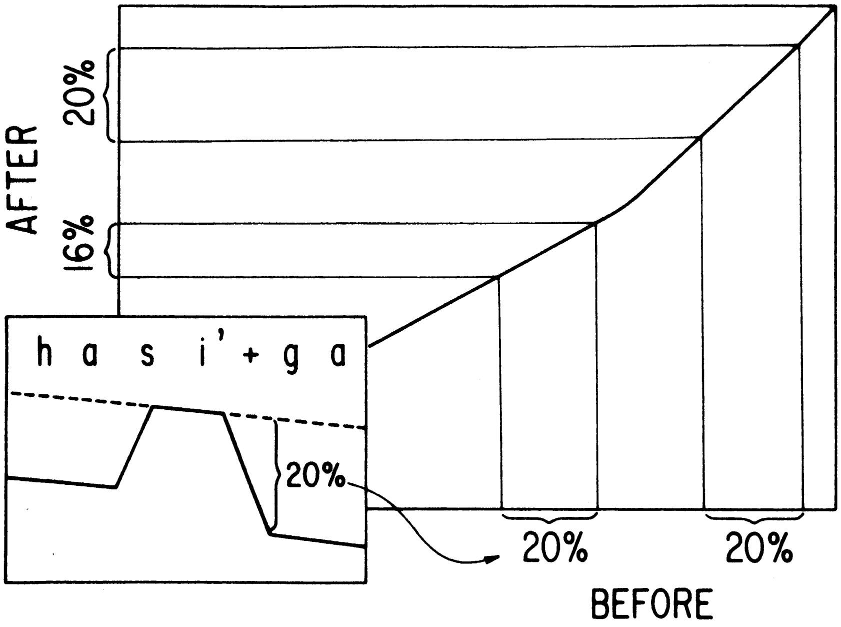

Figure 14. The Pitch-Scale Transform

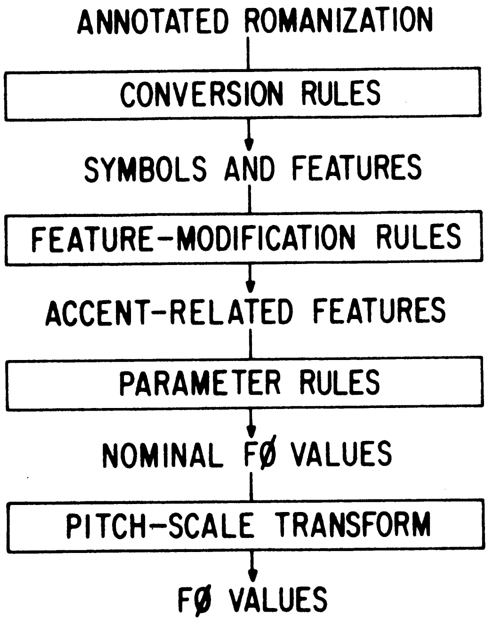

Figure 15. Review

The Pitch-Scale Transform [with examples]

Consider, for example, the accentual drop in hasi'-ga. Before the transform applies, this drop is implemented as a 20% lowering from the original topline value, as illustrated in Figure 14. After the transform, however, because it occurs at the lower frequencies relatively late in the utterance, it is reduced to a fall of only 16%. The effect of the pitch-scale transform is especially evident when the utterance has a long major phrase.

Review

In summary, we will retrace the steps in our synthesis of Japanese pitch patterns (see Figure 15). First, the text of the utterance to be synthesized is annotated with accent marks and boundary symbols. Next, the annotated text is passed to a set of conversion rules, which rewrite the text as a string of symbols and associated features. Then, the utterance is passed to a set of feature-modification rules, which add accent-related features to the utterance segments. These features are then sent to a set of parameter rules, which use the features to produce a set of nominal FØ values for the utterance. Finally, the values generated by the rules are fed to a pitch-scale transform function that produces the final FØ values to be played on the synthesizer. The only step in this process that is not accomplished automatically by our rule system is the initial annotation of the input text that specifies prosodic information, absent in standard romanizations.[9]

----

[9] We have not as yet formally tested how easy it is for users to learn our present annotation system or whether the prosodic units specified are intuitively obvious to native users.

References

Arisaka, Hideyo. 1941. Akusento no kata no honsitu ni tuite. *Gengo kenkyuu* 7: 83-92.

Akinaga, K. 1958. Tokyo akusento no shuutoku hoosoku. In Kindaichi, *Meikai nihongo akusento jiten* (Tokyo: Sanseido).

Fujimura, Osamu. 1966. Kotoba no kagaku (5). *Shizen* 4: 51-57.

Hattori, Shirô. 1947. "Bunsetsu" ni tsuite. (Reprinted in Hattori [1960]).

-----. 1949. "Bunsetsu" to akusento. (Reprinted in Hattori [1960]).

-----. 1954. On'inron kara mita kokugo no akusento. (Reprinted in Hattori [1960]).

-----. 1960. *Gengogaku no hoohoo* (Tokyo: Iwanami).

Hertz, Susan R. 1982. From text to speech with SRS. *J. Acoust. Soc. Am.* 72: 1155-1170.

Kawakami, Shin. 1966. Accentuation of monosyllabic particles preceded by nouns. *Study of sounds* 12: 239-253.

-----. 1977. Size and strength of an accent unit. *Kokugogaku* 111: 78-83.

McCawley, James D. 1968. *The phonological component of a grammar of Japanese* (The Hague: Mouton).

Miyata, K. 1927. New view on the Japanese accent and its notation. *Study of sounds* 2: 18-22.

Miyata, K. 1928. My view on Japanese accent. *Study of sounds* 2: 233-240.

Sagisaka, Yoshinori, and Hirokazu Sato. 1983. Secondary accent analysis in Japanese stem-affix concatenations. Transactions of the Committe on Speech Research, no. S83-05 (Tokyo: The Acoustical Society of Japan).