Affrication as a Performance Device*

Kuniya Nasukawa        Phillip Backley
Tohoku Gakuin University          Tohoku Gakuin University

ABSTRACT. Affricates are traditionally analysed as contour segments containing both values of the feature [±cont], where the sequential ordering of [–cont] and [+cont] must be stipulated in representations to produce the stop-fricative contour. Owing to the problems associated with this view, we argue that affricates should instead be represented phonologically as simplex stops, proposing that affrication is no more than a performance device intended to enhance perception. Using the Element Theory model, we show how stops at certain places of articulation are overly rich in linguistic information and therefore difficult for listeners to recover. By producing these stops with affrication, speakers assist listeners in the perception process.

Keywords: Element Theory, affricate stops, speech perception, place of articulation, acoustic patterns

1. Affricates as Contour Segments

In order to describe contour segments such as affricates, it has traditionally been assumed that precedence relations should be expressed within segments as well as at the skeletal (i.e. CV) level. Sagey (1986), for example, proposes the representation in (1) for affricates, where the ordering of [–cont] and [+cont] reflects the order in which these feature values are phonetically interpreted.

(1) Linear ordering of [–cont] and [+cont] in affricates (Sagey (1986)):

In recent years, however, the grammatical status of contour representations has been called into question by Lombardi (1990), Schafer (1995), Scobbie (1997), Scheer (2003), Nasukawa (2005) and others, who highlight a number of problems associated with contours. First, it is difficult to account for the way in which affricate contours defy typical edge effects (Lombardi (1990)). Second, it cannot be explained why the two slots in a contour never appear in the reverse order, even though this should be at least a possibility: e.g. [dʒ] vs *[ʒd]. And third, although the number of segmental slots in an affricate is always two, there is no immediate reason why this should be so. Indeed, if the phonology allows contours to contain two slots, then why not three or more? This would, of course, multiply the set of possible contour segments far beyond anything we observe in natural languages, but nevertheless the (arbitrary) upper limit of two slots deserves explanation.

2. Affricates as Single Segments

In response to these points, there is now growing support for the view that the precedence relations we find in contour segments are not due to the sequential ordering of feature values in representations. Rather, they are understood to result from the staggered realisation of single segmental structures. For example, Lombardi (1990) proposes a Feature
Geometric representation containing the two unordered privative stricture features [cont] and [stop]; she proposes that these two features reside on separate autosegmental tiers and exhibit a symmetric dependency relation. On the other hand, Shafer (1995) adapts Lombardi’s proposal by introducing an asymmetric dependency relation between [stop] and [cont]. These and other similar proposals all dispense with formal precedence relations between the relevant segmental features. Here, we accept the view that affricates are not contour segments, since there is insufficient phonological evidence to support any segment-internal ordering of features. That is, we recognize no phonological difference between plain stops and affricate stops; affrication itself is taken to be entirely a matter of phonetic realisation.

Yet the question still remains: why are some stops produced as contours (i.e. as affricates, with a staggered realisation), while others are produced as non-contours (i.e. as plain stops, with a simultaneous realisation)? In the remainder of this paper we develop an answer to this question based on speech perception. Although a perceptual account does depart from current thinking on the nature of affricates, we explain below how this angle is particularly well suited to the Element Theory (Harris and Lindsey (1995), Nasukawa and Backley (2005)) approach to segmental representation adopted here. In many ways, Element Theory offers advantages over the use of traditional features for describing segmental structure. For example, it imposes tight control over which expressions, contrasts, and processes the phonology is able to generate. Yet when it comes to explaining affrication, Element Theory initially appears to be at a disadvantage compared with feature-based descriptions, because it has no direct equivalent to the relevant features such as [cont], [strident] or [delayed release]. In fact, nothing in the Element Theory vocabulary points towards a manner difference that would distinguish affricates from plain stops.

This is not necessarily a problem, however. Rather, it compels us to look beyond representations for an explanation of affrication. Our own view is that the appearance of affricates is connected with speech perception: by affricating a stop in situations where perception is difficult, we enhance the recoverability of certain acoustic cues associated with that stop. Before developing this argument further, the following section sets the theoretical context by outlining the reasoning behind Element Theory representations.

3. Element Theory
3.1. Segmental Structure in Element Theory

Element Theory describes segmental structure using the set of six privative features or elements \(\text{\{}\text{A I U H N} \text{\}}\), which are assumed to be present in all languages. The categories defined by the elements are strictly phonological in nature, since they emerge through the observation of phonological phenomena and form the basis of lexical contrasts. As such, elements may be viewed as mental or internal objects containing linguistic information, which serve to distinguish one morpheme from another. But because language functions primarily as a communicative tool, and because communication involves the transfer of
information through some physical medium such as speech, it follows that these internal objects must also make some reference to the external world.

Since the publication of *The Sound Pattern of English* (Chomsky and Halle (1968)), the prevailing view has been to assume that articulation acts as the bridge between the internal and the external — between mental objects and their physical realisation — which accounts for the clear articulatory bias of most segmental descriptions. For the reasons given in Harris and Lindsey (2000), however, Element Theory rejects this speaker-oriented view in favour of an alternative approach in which phonological objects are associated with properties of the acoustic signal. Yet this does not mean that elements are defined in terms of raw acoustic properties such as formant values; rather, elements represent “those information-bearing patterns which humans perceive in speech signals” (Harris and Lindsey (2000), pp. 186-7)).

The Element-based approach assumes that listeners instinctively seek out linguistic information: when decoding speech, they ignore most of the incoming acoustic stream and focus only on the specifically linguistic information contained within the speech signal. Thus Element Theory recognizes the human ability to extract from running speech those acoustic patterns that are relevant to language; and furthermore, it assumes that the mental phonological categories represented by elements are mapped directly on to those same patterns. So, although elements can be physically expressed as such patterns, they exist primarily as mental constructs in the internalized grammar. By describing an element as an ‘auditory image’, Harris and Lindsey (2000) draw attention to this ambiguity: they make it clear that an element is primarily an internal object — a mental image of some linguistically significant information; but secondarily, it is also an external object — a physical pattern in the speech signal which listeners use to cue that mental image.

In principle, any element may occupy any syllabic position. As shown in Nasukawa and Backley (2005), however, the intrinsic properties of an element do affect the likelihood of that element appearing in specific positions. In other words, some elements are more/less marked than others in a given syllabic position: in general, |A I U| tend to appear in nuclear positions, whereas |H N ?| are usually to be found in non-nuclear positions. This distribution is reflected in the division of the element set into two groups: |A I U| form the ‘resonance’ (‘core’) group, while |H N ?| constitute the ‘edge’ (‘peripheral’) group. The distributional characteristics of the elements are shown by their respective positions on the following relational scale indicating nuclear/non-nuclear (core/peripheral) propensity:

\[
\begin{array}{cccc}
|A| & |I| & |U| & |H| & |N| & |?| \\
\hline
\end{array}
\]

Up to this point we have outlined only a listener-oriented view of elements, in which listeners recover the relevant linguistic patterns from the perceived acoustic signal and then associate them with the corresponding element structure. In the case of speakers, the same
linguistic patterns function not as cues but as acoustic targets. Speakers have knowledge of the mapping between the elements in a lexical representation and their associated acoustic patterns; so in order to phonetically interpret a lexical form, speakers must use their vocal apparatus (in whichever way is appropriate) to reproduce the target acoustic patterns themselves. Crucially, this procedure succeeds without the need for an element to contain any articulatory information. In fact, to specify such properties would be counter-productive, as speakers typically have available to them several different articulatory configurations, all of which would serve to produce a similar acoustic result (a ventiloquism act provides the prime example of this).

More relevant to the present discussion is the point that each element has several physical manifestations — that is, the same element may have a different phonetic interpretation depending on its syllabic position, its co-occurring elements, and its status as a head or a non-head (cf. Dependency Phonology, Particle Phonology). For example, when the resonance elements |A I U| appear in a nuclear position, they have the physical attributes shown in (3). It will be recalled from (2) that resonance elements are unmarked in syllable nuclei.

\[(3) \quad \text{The resonance elements in nuclei} \]

<table>
<thead>
<tr>
<th>Elements</th>
<th>Typical acoustic correlates</th>
<th>Typical articulatory execution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mass Central spectral energy mass (convergence of F1 and F2)</td>
<td>Maximal expansion of oral tube; maximal constriction of pharyngeal tube</td>
</tr>
<tr>
<td>[I] dip Low F1 coupled with high spectral peak (convergence of F2 and F3)</td>
<td>Maximal constriction of oral tube; maximal expansion of pharyngeal tube</td>
<td></td>
</tr>
<tr>
<td></td>
<td>rump Low spectral peak (convergence of F1 and F2)</td>
<td>Trade-off between expansion of oral and pharyngeal tubes</td>
</tr>
</tbody>
</table>

In contrast, the edge elements |H N ?| are unmarked in non-nuclear position, and serve primarily to represent distinctions which are typically associated with onsets (and to a lesser extent, codas). In traditional terminology, such distinctions may be described as involving ‘manner’ and ‘laryngeal-source’ properties, as shown in (4):

\[(4) \quad \text{The edge elements in non-nuclei} \]

<table>
<thead>
<tr>
<th>Elements</th>
<th>Typical acoustic correlates</th>
<th>Typical articulatory execution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>edge Abrupt and sustained drop in overall amplitude</td>
<td>Maximal expansion of oral tube; maximal constriction of pharyngeal tube</td>
</tr>
<tr>
<td></td>
<td>low Broad resonance peak at lower end of the frequency range</td>
<td>Velopharyngeal port open</td>
</tr>
<tr>
<td></td>
<td>high Aperiodic energy (noise)</td>
<td>Narrowed stricture producing turbulent airflow</td>
</tr>
</tbody>
</table>

Unlike traditional features such as [+high] or [–son], a single element may stand alone in a segmental expression; for example, in a nucleus the element |A| can be interpreted as the low vowel [a] without needing the support of any additional properties from other elements. However, most segments are represented by compound expressions, which involve a combination of elements interpreted simultaneously. In a compound containing two elements, the phonological characteristics of both elements are expected to emerge in some aspect of the segment’s behaviour. For example, a mid vowel composed of |I A| will usually behave as a front vowel (owing to the presence of |I|) and also as a non-high vowel.
Compounding also affects phonetic interpretation. When two or more elements are interpreted together, the result is an acoustic signal which is often richer in linguistic information because it can contain multiple acoustic patterns. For example, the mid vowel expression /AA/ is interpreted phonetically as [e], its acoustic profile being a combination of two of the patterns given in (3): |AA| contributes ‘mass’ (central spectral energy mass) and |A| provides ‘dip’ (low F1 coupled with high spectral peak). Additionally, some languages add a further level of complexity to the way in which elements combine. This comes about, for instance, when the language in question makes a height distinction within the mid vowel region. For example, some languages contrast [e] versus [e], which are both represented by the same elements /AA/. In this case /A/ and /AA/ may enter into a head-dependency relation such that, when /A/ is headed (i.e. dominant), the whole expression is phonetically mapped onto the close mid vowel [e], since this vowel approximates more closely to the signal pattern for /A/ than for /AA/. Without any dependency relation, /A/ and /AA/ make equal contributions to the compound and the resulting interpretation is [e]. Conversely, when headed by /AA/ the expression is interpreted as the open vowel [æ]. For a similar treatment of edge elements, see Nasukawa and Backley (2005) and Backley and Nasukawa (2007).

Element Theory does not impose any formal restrictions on compounding, so in principle at least, elements are free to co-occur in any combination. However, most languages do show a tendency to prefer some combinations over others. Referring back to (2), we note that /I/ and /UI/ occupy the same position on the nuclear/non-nuclear scale, as do /HI/ and /HI/; and in general, the members of such pairs rarely combine. This means that /AA/ readily combines with either /A/ or /U/, but the expression /AA/ /UI/ is relatively marked (though still possible: e.g. Turkish /AA/ /UI/ = [y/ui]). Below we return to the question of markedness in resonance element combinations when we discuss the representation of consonants.

As already mentioned, elements are not tied to particular syllabic positions, but the same element will be subject to a different phonetic interpretation according to the position where it does appear. As a case in point, the resonance elements /AA/ /I/ /UI/ have the physical values shown in (3) when they appear in nuclear position, whereas in non-nuclear position they contribute consonant ‘place of articulation’ properties. As non-headed elements, they are interpreted in non-nuclear position as follows: /I/ represents the category ‘coronal’ and is typically interpreted as dental; /UI/ represents the category ‘dorsal’ and is interpreted as velar; and /AA/ represents the category ‘guttural’ and is interpreted as a back consonant such as pharyngeal. When the same elements appear in their headed form, they represent the place specifications shown in (5). In each case there is a strong acoustic similarity between the headed and non-headed interpretations of the same element.
(5) Interpretation of headedness in edge elements

<table>
<thead>
<tr>
<th>Elements</th>
<th>Typical articulatory correlates</th>
</tr>
</thead>
<tbody>
<tr>
<td>(</td>
<td>A</td>
</tr>
<tr>
<td>headed (</td>
<td>A</td>
</tr>
<tr>
<td>(</td>
<td>I</td>
</tr>
<tr>
<td>headed (</td>
<td>I</td>
</tr>
<tr>
<td>(</td>
<td>U</td>
</tr>
<tr>
<td>headed (</td>
<td>U</td>
</tr>
</tbody>
</table>

In view of the positional markedness differences shown in (2), it is entirely expected that \(|I|\) and \(|U|\) should appear more frequently than \(|A|\) in non-nuclear positions; and this is indeed cross-linguistically true. A further generalisation emerging from (2) is that all resonance elements are relatively marked in non-nuclear position, and accordingly, it is not surprising to find that most onset and coda (i.e. consonant) expressions contain only a single resonance element. In this paper, however, we examine a particular case where the combining of resonance elements is possible in non-nuclear position. This will provide the foundation for our analysis of affricates, in which we claim that the combination of resonance elements in affricate stops is interpreted quite differently from the way the same elements are realised in nuclear position.

We focus on the role of \(|A|\) in consonants, looking particularly at the way this element contributes to compound resonance. As discussed by Scheer (2003), we assume that \(|A|\) has no salient acoustic pattern to contribute to the overall acoustic profile of a consonant when it co-occurs with another resonance element. Instead, we will claim in the remainder of §3 that \(|A|\) plays the functional role of supporting the consonant’s other ‘place’ properties (represented by either \(|I|\) or \(|U|\)). Effectively, the addition of \(|A|\) increases the structural complexity of an expression without introducing any characteristic signal pattern of its own. By contrast, in those relatively marked cases where \(|A|\) appears as the sole resonance element in a consonant expression, it does contribute a distinct acoustic pattern which is typically interpreted as pharyngeal or epiglottal.

In the analysis below, we focus on unmarked combinations such as \(|I A|\) and \(|U A|\), in order to demonstrate the role of \(|A|\) in non-nuclear positions. Our claim is that these combinations are relevant to the representation of contour segments such as affricates.

### 3.2. Representing Fricative Place

First, using fricatives as an illustration, this section shows how resonance elements define the most common place categories in obstruents. Later we address the issue of stops and affricates. In (6), the single resonance elements in column 1A correspond to the broad consonant categories in 1B. Note how an element can occur as a head (underlined) or as a non-head.
Non-headed elements are mapped onto the broad, unmarked place properties of non-nuclear positions: |I|, |U| and |A| are phonetically interpreted as coronal, dorsal and guttural, respectively. On the other hand, their headed counterparts show a more specific physical manifestation indicating particular place properties: |I|, |U| and |A| are mapped onto palatal, labial and epiglottal, respectively (headed elements are underlined). It is perhaps not immediately clear why the headed and non-headed forms of the element |U| appear to be unrelated in articulatory terms when compared with other resonance elements. In fact, labials and velars are united by similar acoustic properties (principally, a lowering of all formants to yield their characteristically dark sound).

As shown in column 2A in (6), some fricatives require only a single resonance element to describe their place category (i.e. bilabials, dentals, palatals, velars, pharyngeals, epiglottals). For the remaining place categories (labiodental, alveolar, postalveolar, uvular), resonance elements must combine to form complex expressions, as shown in 3A. The mechanism of element combination involves adding |A| to either |I| or |U|, which results in a complex form which is similar to the original simplex form in several respects:

- In representational terms, only the additional |A| separates complex from simplex.
- In acoustic terms, the complex structure has an increased overall intensity — see (7).
- In articulatory terms, the effect of |A| is to cause only a minimal shift in the locus of frication away from the original place of articulation. This is attributed to the intrinsic relational properties among elements: two resonance elements are interpreted as a single segment when they appear in nuclear positions; on the other hand, they receive a staggered interpretation when in most likely positions, non-nuclei.

In (7) the spectral profiles of these fricatives are shown (when produced between low vowels). In each pair, the fricative on the left is simplex (without |A|) while the one on the right is complex (with |A|). It is evident from (7) that by adding |A| we increase the overall intensity of the sound, as indicated by darker shading in the complex expressions. We provide no precise acoustic descriptions here, since the overall impression to be gleaned from the illustrations in (7) is sufficient to demonstrate the role of |A|.
3.3. Representing Stop/Affricate Place

As any typological survey will show, in many languages the set of fricatives is not matched by a corresponding set of homorganic stops; so some fricative categories such as velar have a corresponding stop, whereas others such as labiodental do not. Several authors (Clements (1999), Scheer (2003)) have noted how gaps in the stop system tend to be filled by affricates with the missing place specification in their fricative portion. The table in (8) shows how this pattern of Plosive-Affricate Complementarity (Clements (1999)) produces a stop-fricative pair for every place category (the case of pharyngeals and epiglottals is discussed below).

<table>
<thead>
<tr>
<th></th>
<th>1A</th>
<th>1B</th>
<th>2A</th>
<th>2B</th>
<th>2C</th>
<th>3A</th>
<th>3B</th>
<th>3C</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Labial</td>
<td>I</td>
<td>φ / β</td>
<td>p / b</td>
<td>I</td>
<td>f / v</td>
<td>pf</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>Coronal</td>
<td>II</td>
<td>θ / δ</td>
<td>t / d</td>
<td>II</td>
<td>s / z</td>
<td>ts / dz</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>Palatal</td>
<td>III</td>
<td>ç / ʃ</td>
<td>c / ç</td>
<td>IIAI</td>
<td>ʃ / ʒ</td>
<td>tʃ / dʒ</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>Dorsal</td>
<td>I</td>
<td>x / ɣ</td>
<td>k / ɣ</td>
<td>IIAI</td>
<td>χ / ʁ</td>
<td>kχ</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pharyngeal</td>
<td>I</td>
<td>h / ʃ</td>
<td>—</td>
<td>IIAI</td>
<td>χ / ʁ</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Epiglottal</td>
<td>I</td>
<td>h / ʁ</td>
<td>—</td>
<td>IIAI</td>
<td>χ / ʁ</td>
<td>—</td>
<td></td>
</tr>
</tbody>
</table>
Here we show how Element Theory offers an explanation for why a plain stop appears in some place categories while an affricated stop appears in others. An obvious pattern emerges from (8), such that a stop is affricated whenever it has a complex resonance structure (i.e. two resonance elements). This is, of course, merely a pattern. It does not constitute a reason why two resonance elements in combination should trigger stop affrication, whereas a single element does not. Scheer (2003), working within a variation of the Element Theory model used here, addresses the same issue. He observes that the affricating categories all contain \(|A|\), and by nature this element conflicts with its polar property \(|?|\) (where \(|?|\) is phonetically manifested as the drop in amplitude observed in stops but not fricatives). Scheer claims that this conflict prevents \(|A|\) and \(|?|\) from being interpreted simultaneously, and instead they are realised sequentially as a (phonetic) contour segment.

We agree with Scheer that affrication should be seen as a phonetic device for rescuing phonological structures which are in some way problematic, and allowing them to be interpreted. We also agree that \(|A|\) and \(|?|\) do not naturally co-exist, on account of their polarity, as illustrated in (2). In (8) this is reflected in the absence of pharyngeal stops, which would need to contain the combination \(A ?|\). However, in this paper we argue for an alternative approach to affrication which appeals to speech perception. According to this novel approach, the difference between affricates and plain stops refers not to the presence or absence of \(|A|\) but simply to the number of resonance elements in a stop’s representation: complex resonance (i.e. 2 elements) is interpreted as an affricated stop, while simplex resonance (i.e. one element) results in a plain stop. Below we develop the idea that complex resonance is phonologically grammatical but nevertheless perceptually problematic. This will ultimately lead to the claim that the place of articulation cues associated with complex resonance are more easily recovered in affricated stops than in plain stops.

In order to explain why complex resonance is more difficult to perceive than simplex resonance, let us examine the nature of phonological elements in further detail. As already mentioned in §3.1, an element is primarily a phonological unit, not an acoustic one. It is phonetically mapped onto a corresponding speech signal where we can identify patterns which speakers and listeners use to convey and recover linguistic information about the shape of morphemes (Harris (2006)). For the purposes of lexical access, listeners must identify and extract these linguistically significant patterns from the acoustic signal, then match them with their associated element structures. It therefore seems natural to assume that the greater the pattern complexity of a sound, the more difficult it is for listeners to identify its constituent patterns; this is because a complex sound contains multiple acoustic patterns superimposed on each other. The reader is referred back to the spectrograms in (7), where greater spectral complexity is manifest as a more intense spectral profile.
4. Perception of Place Cues

Having identified a link between element complexity and ease of perception, let us consider further the question of how linguistic information — specifically, consonant place of articulation — is perceived. What are considered to be the most reliable acoustic cues for perceiving place properties? In the case of vowels, identifying place relies primarily on formant frequency: for example, vowel distinctions refer to the pattern behaviour of the first three formants. In the case of obstruents, however, formant structure has at most a secondary role in signaling place, and its relative importance varies from one language to another. Formant transitions in and out of an obstruent do provide some place cues, but the necessary phonological context (i.e. a preceding or following vowel) is not always available to allow those transitions to become manifest.

A more reliable indicator of obstruent place comes from the noise spectrum — from the frequency characteristics of stop bursts in plosives (Malécot (1958), Winitz et al. (1972)) and from the frequency of the spectral peak in the aperiodic noise energy of fricatives (Jongman et al. (1998), Wright et al. (1996), Goodacre and Nakajima (2005)). Now, experimental work has shown that fricatives have more robust place cues than stops, and as a result, are less confusable (Hura et al. (1992)). We suggest two reasons for this. First, fricative cues are segment-internal (i.e. they are contained within the frication noise itself) so they do not rely overly on context; in comparison, stops carry segment-external cues in their release burst and in formant transitions. Second, frication noise — which is rich in the acoustic cues which bear linguistic information — has greater duration than a stop burst, and is therefore more accurately and reliably perceived.

This relates to complex-resonance stops in the following way. We assume that a stop containing two resonance elements has a more complex acoustic pattern — and therefore a more intense acoustic pattern — than a stop with only a single resonance element; the former also contains more linguistic information that must be perceived and decoded by listeners. On this basis, we propose that complex resonance requires a more efficient or robust carrier signal than single resonance, in order to ensure that linguistic information is not lost during transmission. Assuming that the noise spectrum is the most efficient way of transmitting place cues in obstruents, and assuming also that affricates have a sustained period of frication (involving aperiodic noise energy, as in fricatives), it is no coincidence that only stops with complex resonance are systematically affricated.

5. Conclusion

We have proposed that stops containing two resonance elements in their representation are interpreted as affricates, which have a window of audible friction in their fricative portion. This window allows the complex acoustic cues and linguistic information associated with complex resonance to be transmitted by the speaker and recovered by the listener more easily. In this way, affrication is seen as a performance device for improving the perceptibility of complex-resonance stops by making complex place cues more
accessible to listeners. This is achieved by enhancing the portion of the speech signal containing aperiodic noise energy, which is relatively rich in place cues.

In contrast, plain stops with a single resonance element show a less complex and less intense acoustic pattern, which can evidently be recovered from a non-affricated realisation of the stop. In this case, listeners can identify the relevant cues using the stop burst and formant transitions alone. Consequently, plain stops require no lengthened period of frication in order to transmit their place cues; these cues represent the information load of a single resonance element, and as such, are both phonologically and acoustically simpler.

Notes
* An earlier version of this paper was presented at the Phonology Forum 2007 at Sapporo Gakuin University on 27 August 2007. We are indebted to members of the audience for their reactions. This paper has also benefited from insightful comments by two anonymous reviewers. Also we thank the editors for allowing us to exceed the page limit. This research was partially funded by the Japanese government’s Ministry of Education, Culture, Sports, Science and Technology under grants 18520390 and 19520429.
1 The phonology literature provides no satisfactory explanation for why the order [d3] is possible whereas the reverse order [3d] is not. The only point of agreement seems to be the assumption that such an explanation lies outside the phonological domain — that the facts must be constrained by the physiological aspects of human speech production.
2 Despite its appearance, (2) serves a different function from that of the more familiar sonority hierarchy. Whereas sonority controls the distribution of whole segments in syllable structure, the scale in (2) influences the composition of individual expressions by imposing markedness restrictions on element distribution.

References


