MAKING SENSE OF DYNAMIC, NON-SEGMENTAL
PHONETICS

J.K. Local
Department of Language and Linguistic Science, University of York, UK

ABSTRACT
This paper considers some aspects of the interpretation of dynamic approaches to phonetic representation. I argue that the most pressing challenges is that of relating a motivated dynamic, non-segmental phonetics to a phonetics-free phonological analysis.

INTRODUCTION
It may seem somewhat odd that towards the end of the twentieth century phoneticians should, at their international conference, devote a special symposium to dynamic non-segmental phonetics (DNSP). By doing so we might be seen treating the topic as contentious. How could this be? Are we suggesting that it might be possible to talk of a non-dynamic, segmental phonetics? Surely it is the case that from the earliest writings we find references to the continuous, co-ordinated nature of, for instance, articulatory activity in speech production and a concern with how best to represent this complex activity. Moreover, is there not a substantial literature that reports instrumental data and analysis of speech that shows that it is clearly dynamic and non-segmental?

It is certainly the case that linguists and phoneticians have for a long period recognised the inherent multidimensional nature of speech production and that instrumental phonetic investigations in various physical domains have attempted to provide precise details of the dynamics and inter-relatedness of components in this system. (See e.g. Ohala’s vignettes from the history of the phonetic sciences [11].) There is also a general, if tacit, assumption that attention to such details is crucial if we are to gain a full and accurate understanding of the organisation of speech. The hope is that they will account for things that currently present themselves as problematic [2].

Notwithstanding this, I do not think it odd at all to be having a such sympo-
segmental phonological categories and structures. This is reflected, to some extent, in the phonetic terminology they use. Typically it is parametric rather than cross parametric. Thus a Prosodic Analyst is more likely to talk of a phono- logical category being exponed by ‘labiaity, voice and plosion’ than by a ‘voiced bilabial stop’, which suggests \[ \text{labiablity, voice and plosron} \]

Scully provides a clear version of the standard formulation of the ‘problem’: ‘Links are needed to bridge the gap between the analysis of speech as a set of discrete, ordered but durationless linguistic units and analyses of the continuously changing acoustic signals, defined along a time axis.’[15] Saltzman and Munhall [16] offer an equivalent formulation in the articulatory domain in which they ‘attempt to reconcile the linguistic hypothesis that speech involves an underlying sequencing of abstract, discrete, context-independent units, with the empirical observation of continuous, context-dependent interleaving of articulatory movements.’ I will attempt to show that a solution to this problem can be developed by formulating a structured non- segmental phonology and elaborating a compositional phonetic interpretation function.

PHONETICS-PHONOLOGY RELATIONSHIP

I have, in a rather unsolemn manner, reformulated the challenge for a robust DNSP as being concerned with the problem of the relationship between phonetics and phonology. There would appear to be three main ‘solutions’ to this problem: (a) maintain a segmental analysis and propose intermediate levels of representation with sophisticated mapping functions [14]; (b) eliminate the distinction between phonetics and phonology and employ the same categories in both [17, cf also 18] (c) develop a non- segmental phonology with an interpretable dynamic non-segmental phonetics [19, 20, 21].

Other participants in this symposium will be addressing issues arising from (a) and (b). I will restrict myself to a consideration of (c). In doing so I will suggest that ‘the problem’ identified in [15] and [16] above which seems to arise when DNSP confronts a ‘discrete phonology’ is spurious. It arises from a view which, in espousing an intrinsic phonetic interpretation (IPI) hypothesis, misconstrues the timeless, relational nature of phonological representation.

Non-segmental phonology and the IPI hypothesis

Building on the work of Firthan prosodic analysts, colleagues and myself at York have been developing a radical non- segmental model of phonological structure. This model is implemented in the natural-sounding YorkTalk speech generation system [19, 21]. The architecture of this approach is driven from that of Firthan Prosodic Analysis [11, 12]. Prosodic representations are treated as entirely relational. They encode no information about temporal or parametric events. In the York approach the phonological representations are constructed as complex attribute-value structures. The constituents of these structures are unordered, there is no distinguished type of phonological constituent and phonological information is distributed over the entire structure and not concentrated at the terminal nodes. These non-segmental representations make it possible to express phonological contrastivity over any appropriate domain in the structure - at phrase domain, word domain, syllable domain, constituent of syllable (onset, rime etc) for instance. The abstract phonological categories and structures of this model are given temporal and parametric interpretation in terms of a dynamic, non-segmental phonetics. A central aspect of this approach is the rejection of the IPI hypothesis which is propounded in a number of con-temporary ‘non-segmental’ approaches where features in the phonology are deemed to embody a transparent phonetic interpretation - typically cued by the featural name [17, 22].

The position I am outlining does not mean that I see no interesting or ‘explanatory’ links between phonetic phenomena and phonological structures. Rather my claim is that if we wish to develop a sophisticated understanding of the relationships between the meaning systems of a language and make sense of their dynamic exponent in speech, then being forced to provide an explicit statement of the detailed parametric phonetic expenents of phonological structure is an essential prerequisite. The feature labels for phonological units we employ may be given mnemonic labels but their relation to the phon substance need not be simple. Because they are distributed over different parts of the codas structure, their interpretation is essentially polysystemic [11]. For example, the interpretation of the contrast given the feature label [+ nasal] at a rime (see also [23] on the phonetic interpretation of ‘alveolarly and plosion’ in codas of English words). Moreover, the occurrence of the phonologically contrastive feature [+ nasal] at some point in the phonological structure may generalise over many more phonetic parameters than those having to do simply with lowering of the soft palate. (cf [24]).

The consequence of this argument is that nothing at all hangs on the name of a phonological feature provided that the canonical naive view of the relationship between phonological categories and phonetic ones is eschewed. All that the ‘naming of parts’ achieves is some kind of mnemonic short hand. This means that provided the semantics of the phonological categories is explicit and formally stated then it really doesn’t matter what they are called. There are two aspects to specifying the semantics: (1) it is necessary to know how the phonological category(ies) in question relate to other phonological categories - that provide a semantic statement of their place within the phonological systems and structures and (2) it is necessary to provide an explicit statement of the phonetic interpretation of the phonological categories because, in Firthian terms, it ‘renews the connection’ with the dynamic parametric phonetic data [11]. I will develop this position in the following section and show that we can construct a simple phonetic interpretation function which will relate non-segmental structures to a DNSP.
PHONETIC INTERPRETATION OF [ATR] IN KALENJIN

I will now use some data concerning the phonetic characteristics of the [ATR] harmony system in the Kalenjin to motivate an abstract non-segmental phonology and to show how such a phonology can be phonetically interpreted. The broad IPA transcriptions below give an impression of some of the phonetic exponents of [+ATR] in Kalenjin. [+ATR] words are given first for each pair:

1. [kʰɛʔʊm] (to sprinkle)
   [kʰɛʔʊn] (to grow)
2. [kʰɛʔʊlu] (to scrape up)
   [kʰɛʔʊm] (to blow)
3. [kʰɛʔʊ] (to dig up)
   [kʰɛʔʊ] (to dig)
4. [pʰɛː n] (meat)
   [pʰɛːn] (hardship)
5. [l] (far)
   [lʰ] (six)

There are a number of phonetic differences between words in the two categories. These occur not only in vocalic portions but also in the consonantal portions of such words. They include phonatory quality, vocalic and consonantal quality and articulation and durational differences.

Phonatory differences

The two sets of words exhibit different kinds of phonatory activity. Words of the [-ATR] set have audible breath phonation as compared with words in the [+ATR] set. Measurements of the open quotient of the glottal cycle made from electrolaryngographic recordings and inverse filtering reveal (statistically significant) differences that can be taken to confirm breathiness of phonation (larger OQ values are found for [-ATR] words). Spectral characteristics of vocalic portions of the two classes also reveal differences commensurate with breathy versus non-breathy phonation. Examination of voice source measurements also suggest different kinds of laryngeal behaviour in moving from voice to voicelessness in the two sets of words. In [+ATR] voicing dies away slowly and continues at low level. In [-ATR] words, by contrast, voicing drops off rapidly.

Vocalic differences

There are striking auditory differences in vocalic quality between words in the two sets. Vocalic portions in [-ATR] words are noticeably more central (and frequently more open) than those in [+ATR] words. (Note that the open [+ATR] vowel has a back [a] quality in the region of CV5; the open [-ATR] vowel has a front quality in the region of CV4 [a].) These vocalics harmonize with appropriate tokens from the [ATR] set:
   [a] for [+ATR]
   [a] for [-ATR]
   Examination of plots of F1/F2 for tokens of each of the [+ATR] vowel in the data confirms the results of impressionistic listening (for example, [+ATR] vowel show lower F1 values than their congeners [-ATR]).

Consonantal differences

Words of the two categories exhibit different types of stricture and ranges of variation in the consonantal portions. In [+ATR] words we find labial, apical and velar closure with burst release, or with close approximation. In comparable words which are [-ATR] closure with burst release is not found. In such words lax fricative portions occur but do so portions with open approximation. There are also noticeable variations in terms of place of articulation. 'Coronals' in [+ATR] words are exponed with apico-alveolar stricture whereas they may be exponed with either apico-alveolar or dental stricture in [-ATR] words.

Durational differences

Consonantal and vocalic portions are durationally different in [+ATR] words. Typically consonantal portions are shorter in [+ATR] words than they are in [-ATR] words. This is particularly noticeable in the closure and release phases of initial and final plosions. Averages of vocalic duration reveal a tendency for [-ATR] vocoids to be shorter than [+ATR] vocoids. However, [+ATR] words are routinely longer (measured from beginning to end of voicing) than are [-ATR] words.

COMPOSITIONAL PHONETIC INTERPRETATION

[ATR] harmony is canonically the kind of phonological organisation which has been given non-segmental status. Even a hard-core segmentalist would be likely to acknowledge that [ATR] in Kalenjin operates in terms of whole syllable structures. However, it is not immediately clear that such phonological approaches (including eg autosegmental phonology and gestural phonology [16]) could deal in any coherent way with the phonetic interpretation of [ATR] here given the range of different phonetic exponents implicated. It would require a certain amount of ingenuity to postulate a non-segmental [ATR] feature with intrinsic phonetic content and find what there is in common between devoicing of coda approximants, breathy phonation, front or back secondary articulation, consonantal length, particular ranges of consonantal variability and any putative advanced position of the tongue root. Even greater problems might arise in making sense of the ‘counter-intuitive’ phonetic interpretation of the open [+ATR] vowel in the region of [a] and the open [-ATR] vowel in the region of [a].

I suggest that a DNSP interpretation of the abstract phonological relationship designated [+ATR] is more appropriately accomplished with explicit statements of temporal and parametric phonetic exponency for various parts of word and syllable structure. This can be achieved by a compositional phonetic interpretation (CPI) function for partial phonological descriptions [19, 20, 21]. I sketch only the broad outlines of a CPI here.

In the CPI function phonological structures and features are associated with phonetic exponents. The phonetics is the semantics of the phonology [13, 19, 20] (cf [23]). As I indicated earlier, the phonological descriptions being interpreted are here taken to be unordered acyclical graph structures with complex attribute-value node labels. The statement of phonetic exponents in CPI has two formally distinct parts: temporal interpretation and parametric phonetic interpretation. Temporal interpretation establishes timing relationships which hold across constituents of a phonological graph while parametric interpretation instantiates interpreted dynamic ‘parameter strips’ for any given piece of structure (any feature or bundle of features at any particular node in the phonological graph). The resulting ‘parameter strips’ can be considered as sequences of ordered pairs where any pair denotes the value of a particular parameter at a particular (linguistically relevant) time. Thus in the general representation:

(node:partial_phonological_description,
   (Time_start, Time_end), parameter_section)

where the node represents any phonologically relevant contrast domain. The time values may be absolute or relative, fixed or proportional. The precise physical domain of the parameter strips (eg articulatory, acoustic, aerodynamic) is not of immediate relevance here.

The ‘compositional’ part of the interpretation function signifies that the ‘meaning’ of a complex expression is a function of the form and meaning of its parts and the rules whereby the parts are combined [26]. The phonological ‘meaning’ of a syllable equals the ‘meaning’ of its constituents. The compositional principle is instantiated by requiring any given feature or bundle of features at a given place in the phonological structure to only have one possible phonetic interpretation. So for instance, in the present case the words:
   (i) [kʰɛʔʊ] ‘good planters’ and
   (ii) [kʰɛʔʊ] ‘plant!’

can be given the following Firthian like, partial representations:

(i) [ATR+ ] (kοα)
(ii) [ATR- ] (kοα)

Here the syllable-domain [ATR] unit as well as being semantically distinctive serves to integrate the other syllabic material (paradigmatically contrastive units) with consequences phonetic exponency as illustrated above). Given this, then the interpretation of (i) might be written as:

CPI((ATR+: ) (kοα)) = {phonetic exponents of kοα}. A more fully specified representation of (i) might be given as:

[ATR+ ] (kοα) Here the units within the syllable are treated as separate entities or sequences of entities.
The superscript symbols $\kappa$ / $\omega$ placed before the units (k) and (o) serve to indicate onset/time domain contrasts ($\kappa$ 'voicelessness'; $\omega$ 'voice'). Such a representation can be reconstructed as a graph with attribute-value node labels, thus:

\[
\begin{array}{c}
\text{[ATR]}^+ \\
\text{[voi-]} \\
\text{[cnt+}, \text{nas-}, \text{str-}, \text{cnm[cmp+, grw+]]} \\
\end{array}
\]

We have formally tested and verified a CPI for Kalenjin within the YorkTalk declarative speech generation system employing acoustic parameters. Discussion and illustration of this work and quantitative details of the phonetic exponents of [ATR] in Kalenjin are given in Local and Lodge [27].

**CONCLUSION**

Recent phonetic work in laboratories across the world has provided a rich diet of DNSP data. Rather than reviewing this work I have chosen here to concentrate on issues surrounding the interpretation of DNSP data. I have done this because it seems to me that whilst we have seen considerable advances in data collection techniques (eg in the articulatory domain [28]) there has not been a commensurate advance in the linguistic interpretation of that data. By examining a small amount of material from Kalenjin I have tried to motivate the need for a consideration of non-segmental phonological categories in the interpretation of phonetic data. I have suggested that a small step in this direction can be achieved by adopting a non-segmental phonology of the Firthian kind and reject analysis in terms of intrinsic phonetic interpretation. Such a step obliges us to devise an explicit phonetic interpretation function and to explore ways in which DNSP data might relate to abstract non-segmental categories. I think it also moves us towards the 'integrahive phonology' for which Ohala argued so persuasively at ICPHS91 [11].

**REFERENCES**


