Patterns of phonetic contrast: Towards a unified explanatory framework

Björn Lindblom
Stockholm University, Sweden
&
The University of Texas at Austin, U.S.A.
E-mail: lindblom@ling.su.se

ABSTRACT

Stevens makes a convincing case for the non-monotonic nature of the relation between articulation and acoustics and between acoustics and perception. This part of the QT rests on a strong theoretical basis and solid experimental evidence. A second part of the theory claims that the stable regions of the phonetic space correspond to, and define, the universal set of distinctive features as derived from linguists' descriptive analyses of numerous languages. This part of the framework casts a wider net and is the main focus of the present commentary.

1. SUMMARY

Since, in the languages of the world, some sound attributes are, and some are clearly not, quantal, there is need for a framework that integrates the strengths of the Quantal Theory (QT) and related approaches into a broader theory. As a contribution towards that goal, we here present a revisit to the structure of phonetic systems. A comparison is made of the QT and the theory of Adaptive Dispersion (TAD) focusing on the treatment of vowels. The two accounts differ in their choice of explanatory criteria: *contrast* vs *stability*. The key result that enables us to propose a unified account, comes from some recent vowel inventory simulations using a psycho-acoustically and physiologically motivated auditory model. Driving the vowel inventory predictions by means of a measure of contrast defined at the systemic level, we found [1] that, in many, but not all, cases, the simulations had the effect of 'seeking out' quantal points (i.e. salient and stable points according to QT terminology) in the vowel space. QT and TAD converged on similar choices. This result can be attributed to the use of a model that incorporates both spectral and temporal processing. In this type of model, there is no explicit 'formant tracking', but spectral prominences (formant peaks, and 'quantal' regions) are nevertheless accorded a special status. This finding is a compelling illustration of the point Stevens has often made over the years about the non-monotonicity of the phonetic space.

2. WHAT THE QUANTAL THEORY DOES

The Quantal Theory does two important things: (1) First, it demonstrates how articulatory processes map onto acoustics and how acoustic patterns are transformed into perceptual representations. The QT states that this mapping is *non-monotonic,* i.e., the acoustic consequences of equal steps along an articulatory dimension are *quantal*. Stevens illustrates this claim with the well-known schematic where continuous articulatory variation produces three regions in the acoustic response: first a plateau with little change, then a region with rapid change, and then another stable plateau.

(2) Second, it formulates a hypothesis about the origin of 'distinctive features', the phonetic attributes that languages are built from. In short, this hypothesis says that there are regions in phonetic space that are stable, that those regions are functionally highly valued and that it is those regions that languages 'seek out' in selecting their sound systems.

3. SIGNIFICANCE

The historic importance of the QT for our field is beyond all doubt. By building on his insights into acoustics Stevens has comprehensively documented the non-uniform nature of articulatory-acoustic-perceptual mapping. He has also given a characterization of the relation between speech signals and the underlying sound structure that has broad implications for virtually all aspects of phonetic research. It is a theory rich in consequences for accounts of how speech evolved both in the species and over historical time, how it is developed and how it is used in on-line interactions between adult speakers and listeners.

An additional significant aspect of the theory is the fact that it touches on one of the most fundamental attributes of human language. By proposing a phonetic basis for distinctive features – the ultimate building blocks of speech patterns-, it takes steps towards answering why, at two levels of structure – phonology and syntax - languages make combinatorial use of discrete units in organizing vocabularies and building utterances. The universal here referred to is related to Martinet's '*la double articulation*' [2] and Hockett's '*duality of patterning*' [3]. It is part of what Chomsky currently calls the '*discrete infinity*' of language [4], and Studdert-Kennedy attributes to the so-called *particulate principle* [5].

Can the causes of this profound design principle be explained and understood? Or must it be treated merely as a formal idiosyncracy? The QT presents a preliminary

ISBN 1-876346-48-5 © 2003 UAB

phonetically based, explanatory account.

4. FOCUS OF THIS COMMENTARY

When it comes to the two parts of the QT, the first - dealing with the relations between the articulatory, acoustic and perceptual levels - should inspire little controversy. Stevens has conclusively demonstrated the non-monotonicity of the phonetic ‘search space’ both by means of solid empirical evidence and with quantitative theoretical arguments. However, it is when Stevens moves on to speculate on how that space is used by the languages of the world that a fair number of his colleagues may feel provoked to voice their skepticism. My commentary will focus on this part of the QT.

5. CONTRAST OR STABILITY?

To advance the debate about QT and distinctive features and to highlight some of the main issues we will focus on vowels.

According to QT, the point vowels [i a u] are favored because they are quantal. For instance, the F1 and F2 values of [a] form a proximity region. According acoustic theory, in particular the rules governing the predictability of formant levels, it is the case that, when two formants come close together, the amplitude of both increases. The F-pattern of [a] is also relatively insensitive to articulatory imprecision. Accordingly, [a] and certain other vowels are highly valued because of their *salience* and *stability*. According to the theory of adaptive dispersion (TAD) [6], vowels including the point vowels are selected so as to form a system of units that exhibit sufficient perceptual *contrast*.

These two position are linked to different views of how on-line speech communication works.

Proponents of QT assume that perceptual processes extract invariant properties from the speech wave. Consequently, sound systems with acoustically stable units should be favored.

Proponents of TAD argue that the demand for perceptual contrast springs from a condition that all functionally adequate lexical systems must meet: *Distinct meanings must sound different*. Phoneticians holding this view also suppose that the acoustic signal is not *by itself* responsible for successful lexical access but interacts with listener knowledge. The speech percept is never a raw record of the signal. Hence, signal invariance is not a necessary condition for successful speaker-listener interaction [7,8].

Can those positions be reconciled? As indicated below, there is reason to believe that, although they put emphasis on different criteria, it should nonetheless be possible to place them within the same single explanatory framework.

6. VOWEL SYSTEMS REVISITED

In recent work on predicting vowel systems, the effect of noise has been examined. It appears justified to assume that normally speech sounds are not used under perfectly noise free conditions, but typically occur with a certain amount of background noise. Hence it is not unreasonable to expect that phonological systems might show signs of having adapted to those conditions.

This project drew our attention to physiological studies revealing the importance of temporal coding in auditory processing [9] and its role in making signals robust under noisy conditions. The key to noise resistance lies in the fact that strong spectral components (e.g., formant peaks) are carried not only by neural channels with characteristic frequencies (CFs) closest to those prominences but are also redundantly specified across adjacent channels with somewhat different CFs (Fig 1 of [1]).

To improve the auditory realism of vowel specifications we developed a numerical model (KONVERT) that computes auditory spectra using ERB filters as well as Dominant Frequency representations [10]. The Dominant Frequency (DF) of a given auditory filter is calculated from the zero crossings of the output waveform of that filter. Since the DF transform emphasizes strong at the expense of weak components, low amplitudes do not contribute significantly to perceptual distances. As a result, the vowel space is warped relative to a formant-based calibration. Interestingly, the DF-based transform stretches contrasts along F1 (sonority dimension) and compresses contrasts along F2 (chromaticity dimension) (Fig 3 in [1]).

Vowel systems currently predicted with the aid of the dispersion criterion are in excellent agreement with observed facts. Notably the previous problem of ‘too many high vowels’ [1] is eliminated. In previous attempts [11], the predicted seven-vowel system had two high vowels in between [i] and [u] whereas the typical pattern observed in the world’s languages shows [i e ɛ a ɔ o u] as in the systems of Italian and Yoruba (Fig 2 in [1]).

From a purely acoustic point of view the seven-vowel system is a rather remarkable pattern. However, the favoring of the open-close over the front-back dimension is a general characteristic of the typology of vowel phonetics. It is evident not only in vowel systems. We see it in historical vowel shifts, in tense-lax alternations, and in diphthongs [6]. If, for the sake of argument, we assume a vowel space with a range of possible F1 values of 250 to 750 Hz and F2 between 750 and 2250 Hz, we are struck by the fact that F1 shows a much smaller range than F2. Using mel or log frequency units does not eliminate the discrepancy. Given these considerations, one might expect typological data to show fewer vowel distinctions along open-close than along front-back. Rather the reverse pattern is what is actually observed.

To summarize, we find that, when a realistic auditory representation of vowels is used, spectral prominences (formant peaks, and 'quantal' regions) are highlighted. Significantly, the distances between vowels are modified relative to their acoustic distances. Highly relevant to the present topic is the observation that driving the vowel inventory predictions from a distinctiveness criterion - sometimes, but not always - has the effect of 'seeking out' quantal (i.e. salient and stable) points in the space. In that sense, it happens that QT and TAD converge on similar choices, *viz.*, the 'quantal' vowels.

However, they also show divergences. For TAD 'distinctiveness' is the driving force and 'salience' and 'stability' arise as by-products of articulatory-to-acoustic mapping (predictability of formant levels) and acoustic-to-auditory mapping (auditory enhancement of spectral prominences by phase locking). In other words, 'salience' and 'stability' are undeniable properties of some of the selected elements but they play no primary causal roles. TAD is compatible with systems having both 'quantal' and 'non-quantal' phonetic attributes.

There are numerous vowel qualities that occur in the world's languages but that cannot be seen as quantal, e.g., [æ] and [ə]. Such cases pose a problem for the QT program which is committed to finding quantal correlates for all dimensions of phonetic contrast.

> *" …the properties of the sound that is generated by these movements and the human responses to this sound tend to fall into discrete categories. These categories form the bases for the distinctive features that potentially define the phonological contrasts in any language."* (Stevens, abstract, present *Proceedings*).

By comparison, TAD avoids that problem by making the system rather than the individual entities its focus. It is the system of elements that is the object of optimization. A selected unit is highly valued, not because of its individual qualities, but depending on its contribution as a 'team player'. That is the significance of the optimization criterion of the vowel system predictions:

$$minimize \;\; 1/(D_{ij})^2$$

where D_{ij} is the auditory distance between vowels i and j. What is the source of systemic organization? It appears possible to attribute it to perception. This process can be seen as systemic if we assume that, in lexical access, a phonetic signal at first excites the lexical system as a whole but ends up activating only one or a few candidates for further processing.

7. FINAL COMMENTS

The QT inspires us to think about the phonetic and other sources of distinctive features. In so doing it draws attention to the fundamental fact about how human languages are organized: its combinatorial use of discrete units (cf comments above). Languages form words by combining distinctive features into segments, segments into syllables and syllables into words. Combinatorial coding extends to morphology and syntax. It is pervasive throughout linguistic structure. It is what makes human language unique in the animal world.

Imagine a hypothetical code using 25 entities. Suppose 'words' are formed by concatenating 1, 2, 3, 4 or 5 units: x, xx, xxx, xxxx, xxxxx where x = one of the 25 entities. How many distinct words would we be able to make with this code? Answer: $2^5 + 25^2 + 25^3 + 25^4 + 25^5 = > 10^7$. A very large number. Although languages do not exploit this organization fully, it is the key to its expressive power of language.

The functional significance of this coding is, among other things, that is saves memory space. The reuse of a small number of distinctive features, segments etc. can be compared to the formats modern technology uses to compress image and other data files. Metaphorically the combinatorial reuse of linguistic constructs is like a biological form of self-compressed formatting.

Can the causes of this design principle be explained and understood? Or must it be treated merely as a formal idiosyncracy?

In view of its functional significance, it would seem that its origins might have something to do with memory constraints. Perhaps, as suggested by some research [12, 13], the encoding of memory traces is associated with a bio-chemical cost in that it takes time and metabolical energy to build and to use such traces. If so, a criterion of minimum time and minimum energy consumption would, in automatic and unsupervised ways, bias the learning of new phonetic patterns in such a way as to favor shapes that put minimal demands on time and energy. What would that bias entail? It could conceivably induce patterns of reuse and favor forms that are perceptually distinct but nevertheless share some of the articulatory and auditory specifications. In other words, the memory constraint just described would create a mechanism operating as a minimal-pair machine [14]. A possible source of linguistic re-use?

Such ideas are admittedly speculative but serve the purpose here of widening the phonetically based QT perspective. Combinatorial coding of units is not limited to the level of distinctive features. Explaining the origin of distinctive features will undoubtedly need to be integrated with accounting for this organization also at other levels of linguistic structure. Accordingly it will need to go beyond phonetics in the narrow sense – which points our inquiry in the direction of memory and learning.

ACKONOWLEDGEMENTS

The author thanks Randy Diehl and Carl Creeger for their contributions to the research reported here.

REFERENCES

[1] R Diehl, B Lindblom and C Creeger, “Increasing Realism of Auditory Representations Yields Further Insights into Vowel Phonetics”, poster at ICPhS 15th, present proceedings, 2003.

[2] Martinet A, *Économie des changements linguistiques*, Francke, Bern, 1955.

[3] C F Hockett, *A manual of phonology*, Bloomington, Indiana, Indiana University Press, 1955.

[4] M D Hauser, N Chomsky and W T Fitch, “The faculty of language: What Is It, Who Has It, How Did It Evolve?”, *Science*, vol. 298, pp. 1569–1579, 2003.

[5] W L Abler, “On the particulate principle of self-diversifying systems”, *J Social Biol. Struct*, vol. 11, pp. 264–277, 1988.

[6] R L Diehl and B Lindblom: “Explaining the structure of feature and phoneme inventories”, in *Speech processing in the auditory system*, S Greenberg, W A Ainsworth, A Popper & R Fay (eds): New York:Springer Verlag, (in press).

[7] B Lindblom, "Explaining phonetic variation: A sketch of the H&H theory", in *Speech Production and Speech Modeling*, W Hardcastle & A Marchal (eds), Dordrecht:Kluwer, 403-439, 1990.

[8] Lindblom B, "Role of articulation in speech perception: Clues from production", *J Acoust Soc Am* 99(3), 1683-1692, 1996.

[9] B. Delgutte and N. Kiang, “Speech coding in the auditory nerve I: Vowel-like sounds, “*Journal of the Acoustical Society of America*, Vol. 75, pp. 866-878, 1984.

[10]R. Carlson and B. Granström, “Towards an auditory spectrograph,” in *The Representation of Speech in the Peripheral Auditory System*, R. Carlson and B Granström, Eds, pp. 109-114, 1982.

[11]J. Liljencrants and B. Lindblom, “Numerical simulation of vowel quality systems: The role of perceptual contrast,” *Language*, vol. 48, pp. 839-862, 1972.

[12]F Gonzales-Lima, “Brain imaging of auditory learning functions in rats: Studies with fluorodeoxyglucose autoradiography and cytochrome oxidase histochemistry”, in *Advances in metabolic mapping techniques for brain imaging of behavioral and learning functions,* NATO ASI Series D:68, F Gonzales-Lima, T Finkenstädt T & H Sheich (eds), pp 39-109, Dordrecht: Kluwer, 1992.

[13]M T T Wong-Riley, ”Cytochrome oxidase: An endogenous metabolic marker for neuronal activity”, *Trends Neurosci* 12(3), pp 94-101, 1989.

[14] B Lindblom, “Developmental origins of adult phonology: The interplay between phonetic emergents and evolutionary adaptations”, *Phonetica*, K Kohler, R Diehl, O Engstrand & J Kingston (eds), 2000.