Advances in Complex Systems, Vol. 11, No. 3 (2008) 371–392

REDISCOVERING THE CO-OCCURRENCE PRINCIPLES OF VOWEL INVENTORIES: A COMPLEX NETWORK APPROACH

ANIMESH MUKHERJEE, MONOJIT CHOUDHURY, ANUPAM BASU and NILOY GANGULY

Department of Computer Science and Engineering,
Indian Institute of Technology,
Kharagpur, India 721302

SHAMIK ROY CHOWDHURY

Department of Information Technology,
Heritage Institute of Technology,
Kolkata, India 700107

Received 16 April 2007
Revised 22 November 2007

In this work, we attempt to capture patterns of co-occurrence across vowel systems and at the same time figure out the nature of the force leading to the emergence of such patterns. For this purpose, we define a weighted network where the vowels are the nodes and an edge between two nodes (read vowels) signifies their co-occurrence likelihood over the vowel inventories. Through this network we identify communities of vowels, which essentially reflect their patterns of co-occurrence across languages. We observe that in the communities formed by the frequently occurring vowels, the constituent nodes are largely uncorrelated in terms of their features and show that they are formed based on the principle of maximal perceptual contrast. However, in the rest of the communities, strong correlations are reflected among the constituent vowels with respect to their features, indicating that it is the principle of feature economy that binds them together. We validate the above observations by proposing a quantitative measure of perceptual contrast as well as feature economy and subsequently comparing the results obtained due to these quantifications with those where we assume that the vowel inventories had evolved just by chance.

Keywords: Vowels; complex network; community structure; feature economy; feature entropy.

1. Introduction

Linguistic research has documented a wide range of regularities across the sound systems of the world's languages [2, 5, 14, 15, 21–23]. Functional phonologists argue that such regularities are the consequences of certain general principles, like *maximal perceptual contrast*[a] [14], *ease of articulation*[b] [2, 16] and *ease of learnability*[c] [2]. In the study of vowel systems the optimizing principle, which has a long tradition [10, 30] in linguistics, is maximal perceptual contrast. A number of numerical studies based on this principle have been reported in the literature [14, 15, 26]. Of late, there have been some attempts to explain the vowel systems through multiagent simulations [2] and genetic algorithms [11]; all of these experiments also use the principle of perceptual contrast for optimization purposes.

An exception to the above trend is a school of linguists [3, 7] who argue that perceptual-contrast-based theories fail to account for certain fundamental aspects, such as the patterns of co-occurrence of vowels, based on similar acoustic/articulatory *features*[d] observed across the vowel inventories. Instead, they posit that the observed patterns, especially found in larger inventories [3], can be explained only through the principle of *feature economy* [8, 18]. According to this principle, languages tend to maximize the combinatorial possibilities of a few distinctive features to generate a large number of sounds.

The aforementioned ideas can possibly be linked together through the example in Fig. 1. As shown, the bottom plane P constitutes a set of three very frequently occurring vowels, /i/, /a/ and /u/, which usually make up the smaller inventories and do not have any single feature in common. Thus, smaller inventories are quite likely to have vowels that exhibit a large extent of contrast in their constituent features. However, in bigger inventories, members from the higher planes (P' and P'') are also present and they in turn exhibit feature economy. For instance, in the plane P' constituting the set of vowels /ĩ/, /ã/ and /ũ/, we find a nasal modification applied equally on all the three members of the set. This is actually indicative of an economic behavior that the larger inventories show while choosing a new feature in

[a]Maximal perceptual contrast refers to the principle that the phonemes as well as the other linguistic units (e.g. syllables and words) of a language should be maximally distinct from each other, because this facilitates proper perception of the individual linguistic units in a noisy environment.
[b]The principle of ease of articulation states that the structure of a language should facilitate expresssion and dissemination of information with minimal energy spent on the part of the speaker. Two of the general implications of this principle are: frequent words are shorter; the sound systems of all languages are formed from certain universal (and highly frequent) sounds that do not involve complicated articulatory gestures.
[c]The principle of ease of learnability states that a language should be highly learnable in order to propagate through the generations. Consequences of this principle include facts such as: linguistic structures are usually regular; irregularities, if any, are featured only by extremely infrequent linguistic units.
[d]In phonology, features are the elements, which distinguish one phoneme from another. The features that describe the vowels can be broadly divided into three different classes, namely the *height*, the *backness* and the *roundedness*. Height refers to the vertical position of the tongue relative to either the roof of the mouth or the aperture of the jaw. Backness refers to the horizontal tongue position during the articulation of a vowel relative to the back of the mouth. Roundedness refers to whether the lips are rounded or not during the articulation of a vowel. There are, however, still more possible features of vowel quality, such as the velum position (e.g. nasality), the type of vocal fold vibration (i.e. phonation), and the tongue root position (i.e. the secondary place of articulation).

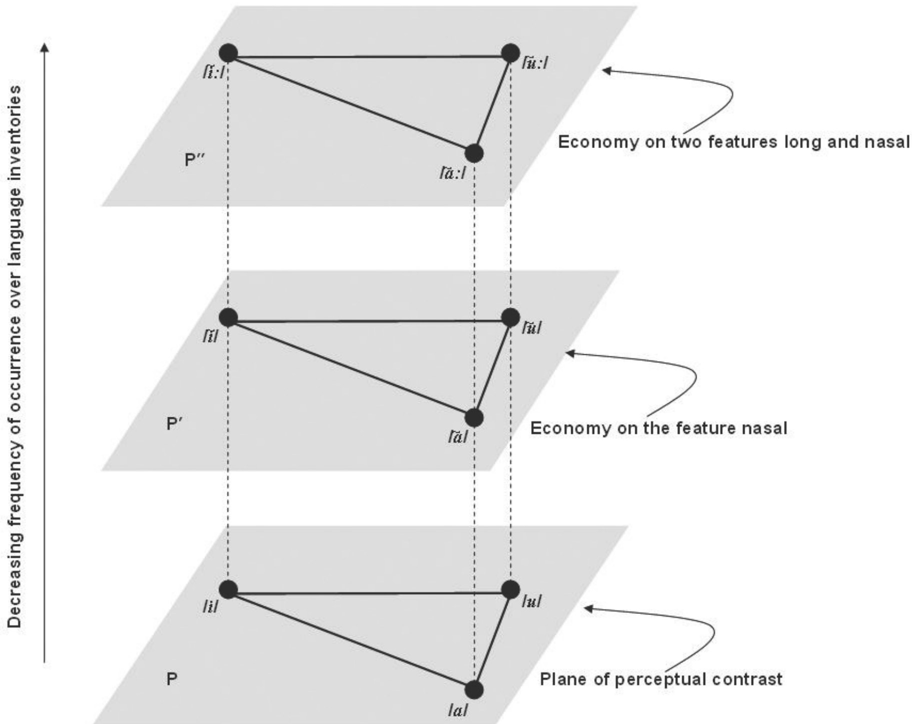


Fig. 1. The organizational principles of the vowels (in decreasing frequency of occurrence) indicated through different hypothetical planes.

order to reduce the learnability effort of the speakers. The third plane P'' reinforces this idea by showing that the larger the size of the inventories, the greater the urge for this economy in the choice of new features. The figure also illustrates another interesting relationship that exists between the vowels across the planes (indicated by the broken lines). All these relations are representative of a common linguistic concept of markedness [7] in which one less frequently occurring vowel (say, /ĩ/) implies the presence of the other (and not vice versa) frequently occurring vowel (say, /i/) in a language inventory. In this co-occurring pair (/i/ and /ĩ/), the frequently occurring vowel (i.e. /i/) is usually referred to as the unmarked member, while the less frequent one (i.e. /ĩ/) is called the marked member. Note that these cross-planar relations are also indicative of feature economy because all the features present in the frequent vowel are also shared by the less frequent one. In summary, while the basis of organization of the vowel inventories is perceptual contrast as indicated by the plane P in Fig. 1, economic modifications of the perceptually distinct vowels take place with the increase in the inventory size (as indicated by the planes P' and P'' in the figure).

In this work, we attempt to corroborate the above conjecture by automatically capturing the patterns of co-occurrence that are prevalent *in* and *across* the planes illustrated in Fig. 1. For this purpose we define the "Vowel–Vowel Network," or VoNet, which is a weighted network where the vowels are the nodes and an edge between two nodes (read vowels) signifies their co-occurrence likelihood over the vowel inventories. In order to capture the patterns of co-occurrence, we conduct community structure analysis of VoNet in and across the planes P, P' and P'' shown in Fig. 1. We also present a quantitative measure for estimating the extent of feature economy for a given set of vowels and show that while the vowel communities in a plane exhibit lower feature economy, the communities across the planes display much higher feature economy. These findings, in turn, imply that the bigger vowel inventories are formed on the basis of feature economy, while the smaller ones are governed by the principle of maximal perceptual contrast. We compare the extent of feature economy observed in real inventories to that in randomly generated inventories, which further corroborates the above findings.

This article is organized as follows. Section 2 describes the experimental setup in order to explore the co-occurrence principles of the vowel inventories. In this section we formally define VoNet, outline its construction procedure, present a community-finding algorithm, and also present a quantitative definition for maximal perceptual contrast as well as feature economy. In Sec. 3 we report the experiments performed to obtain the community structures, which are representative of the co-occurrence patterns in and across the planes discussed above. We also report results where we measure the driving forces that lead to the emergence of such patterns and show that the real inventories are substantially better in terms of this measure than those where the inventories are assumed to have evolved by chance. Finally, we conclude in Sec. 4 by summarizing our contributions, pointing out some of the implications of the current work and indicating the possible future directions.

2. Experimental Setup

In this section, we systematically develop the experimental setup in order to investigate the co-occurrence principles of the vowel inventories. For this purpose, we formally define VoNet, outline its construction procedure, describe a method to extract the community structures from VoNet, and define metrics that can quantitatively capture the co-occurrence principles of the vowels forming these communities.

2.1. *Definition and construction of VoNet*

2.1.1. *Definition of VoNet*

We define VoNet as a network of vowels, represented as $G = \langle V_V, E \rangle$, where V_V is the set of nodes labeled by the vowels and E is the set of edges occurring in VoNet. There is an edge $e \in E$ between two nodes, if and only if there exists one or more languages where the nodes (read vowels) co-occur. The weight of the edge e (also *edge weight*) is the number of languages in which the vowels connected by e co-occur. The weight of a node u (also *node weight*) is the number of languages in

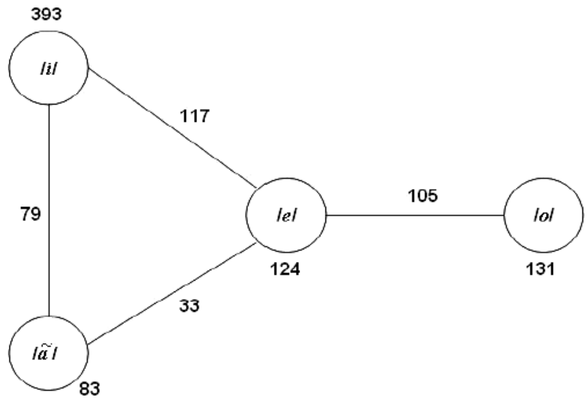


Fig. 2. A partial illustration of the nodes and edges in VoNet. The labels of the nodes denote the vowels represented in the IPA (International Phonetic Alphabet). The numerical values against the edges and nodes represent their corresponding weights. For example, /i/ occurs in 393 languages; /e/ occurs in 124 languages while they co-occur in 117 languages.

which the vowel represented by u occurs. In other words, if the vowel v_i, represented by the node u, occurs in the inventory of n languages, then the node weight of u is assigned the value n. Also, if the vowel v_j is represented by the node v and there are w languages in which the vowels v_i and v_j occur together, then the weight of the edge connecting u and v is assigned the value w. Figure 2 illustrates this structure by reproducing some of the nodes and edges of VoNet.

2.1.2. *Construction of VoNet*

Many typological studies [5, 9, 13, 16, 21, 23] of segmental inventories have been carried out in the past on the UCLA Phonological Segment Inventory Database (UPSID) [17]. Currently UPSID records the sound inventories of 451 languages, covering all the major language families of the world. In this work we have therefore used UPSID comprising these 451 languages and 180 vowels found across them for constructing VoNet. Consequently, the set V_V comprises 180 elements (nodes) and the set E comprises of 3135 elements (edges). Figure 3 presents a partial illustration of VoNet as constructed from UPSID.

2.2. *Finding community structures*

We attempt to identify the communities appearing in VoNet by a modified version of the Radicchi *et al.* [25] algorithm for weighted networks as introduced by us in an earlier article [21]. The basic idea of this modified algorithm (henceforth termed MRad) is that if the weights on the edges forming a triangle (loops of length 3) are comparable, then the group of vowels represented by this triangle frequently occur together rendering a pattern of co-occurrence; while if these weights are not comparable, then there is no such pattern. In order to capture this property we

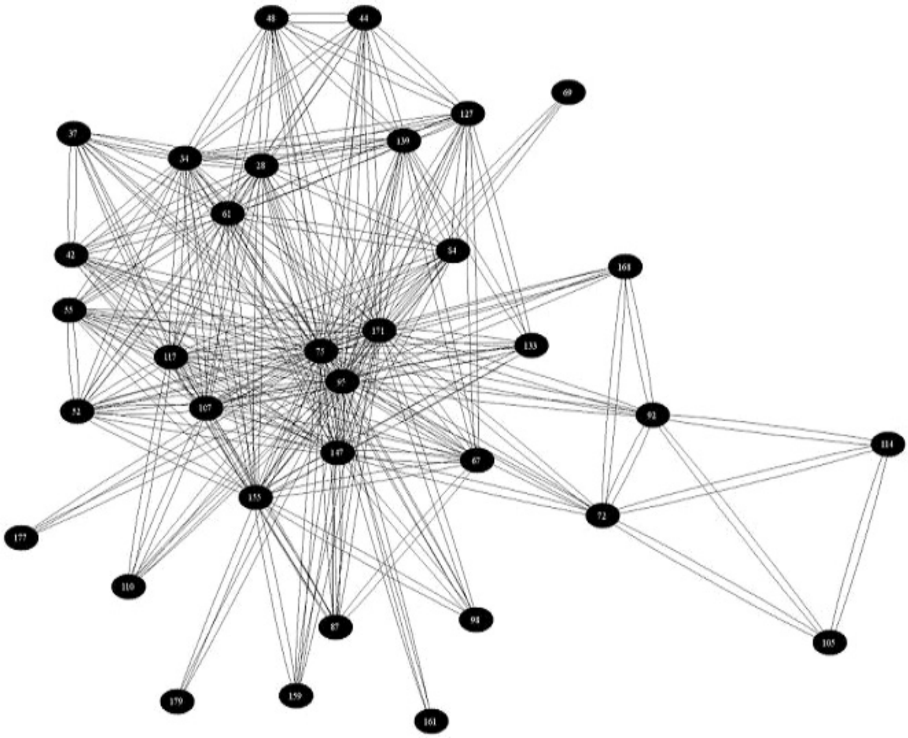


Fig. 3. A partial illustration of VoNet. All edges in this figure have an edge weight greater than or equal to 15. The number on each node corresponds to a particular vowel. For instance, node No. 72 corresponds to /ĩ/.

define a strength metric S for each of the edges of VoNet, as follows. Let the weight of the edge (u,v), where u, $v \in V_V$, be denoted by w_{uv}. We define S as

$$S = \frac{w_{uv}}{\sqrt{\sum_{i \in V_V - \{u,v\}} (w_{ui} - w_{vi})^2}} \tag{1}$$

if $\sqrt{\sum_{i \in V_V - \{u,v\}} (w_{ui} - w_{vi})^2} > 0$, else $S = \infty$. The denominator in this expression essentially tries to capture whether or not the weights on the edges forming triangles are comparable (the higher the value of S, the more comparable the weights). The network can then be decomposed into clusters or communities by removing edges that have S less than a specified threshold (say, η).

At this point, it is worthwhile to clarify the significance of a vowel community. A community of vowels actually refers to a set of vowels which occur together in the language inventories very frequently. In other words, there is a higher-than-expected probability of finding a vowel v in an inventory which already hosts the other members of the community to which v belongs. For instance, if /i/, /a/ and /u/ form a vowel community and if /i/ and /a/ are present in any inventory, then there is a very high chance that the third member, /u/, is also present in the inventory.

2.3. *Definition of the metrics*

After extracting the vowel communities from VoNet by the application of MRad, we investigate the driving force that leads to the emergence of these communities. For this purpose, we propose a quantitative measure that captures both perceptual contrast and feature economy. In order to establish that the above forces really drive the formation of the communities, we also need to compare and show that the communities are much better in terms of this measure than they would have been if the vowel inventories had evolved by chance. In the rest of this section we detail the metric that is used for the purpose of quantification as well as the metric that is used for the purpose of comparison.

2.3.1. *Metric for quantification*

For the characterization of feature economy, we shall use a metric called *feature entropy*, which has been introduced and explained in Refs. 21 and 22. For the sake of readability, we motivate and explain the concept of feature entropy below. For a community C of size N (i.e. comprising N nodes), let there be p_f vowels, which have a particular feature f (f is assumed to be Boolean in nature)[e] in common and q_f other vowels, which lack f. Hence, $p_f + q_f = N$. It immediately follows that the probability that a particular vowel chosen uniformly at random from C has f is $\frac{p_f}{N}$ and the probability that the vowel lacks f is $\frac{q_f}{N}$ $(= 1 - \frac{p_f}{N})$. One can think of f as an independent random variable, which can take values 1 and 0, and $\frac{p_f}{N}$ and $\frac{q_f}{N}$ define the probability distribution of f. Thus, for any given community C, the amount of information (or disorderedness) expressed through the feature f is given by the binary entropy H_f, defined as [27]

$$H_f = -\frac{p_f}{N}\log_2\frac{p_f}{N} - \frac{q_f}{N}\log_2\frac{q_f}{N}. \tag{2}$$

Note that H_f can vary from 0 (when all the members in C have the same value for f) to 1 (when $\frac{p_f}{N} = \frac{q_f}{N} = 0.5$). Since the principle of feature economy predicts that all the vowels will have similar features, a smaller value of H_f implies that a vowel community makes economic use of the feature f. Let F be the set of all the features present in the vowels in C. We define the term *feature entropy*, F_E, as the sum of the binary entropies with respect to all the features:

$$F_E = \sum_{f\in F} H_f = \sum_{f\in F}\left(-\frac{p_f}{N}\log_2\frac{p_f}{N} - \frac{q_f}{N}\log_2\frac{q_f}{N}\right). \tag{3}$$

This method of combining binary entropies of the parts of a system in order to obtain the total entropy of the system is quite prevalent in the literature and is

[e]There are 28 such Boolean features that are found across the vowel systems recorded in UPSID.

often termed *joint entropy* [12, 20, 31]. If the parts of the system are assumed to be independent of one another, then from the subadditivity property [12] the joint entropy of the system becomes the sum of the binary entropies. In the current work we also treat each feature f as an independent random variable, which means that our feature entropy estimate (which follows from the subadditivity property) is an upper bound on the true feature entropy of a community of vowels.[f] However, there are also some works where the features are assumed to be dependent on one another (see Ref. 19, and we plan to extend our work along these lines in future).

It can be shown that F_E is the number of bits that are required to communicate the information about the entire community C through a channel (see Refs. 21 and 22 for explanation). The value of F_E for any community C comprising vowels can range from $\log_2 N$ (under the assumption that each vowel has to be uniquely represented) to $|F|$. C exhibits the most economic utilization of the features if F_E is $\log_2 N$. Stated differently, the higher the value of F_E, the weaker the role of feature economy in C.

2.3.1.1. Capturing perceptual contrast

If C constitutes a set of perceptually distinct vowels, then a larger number of bits should be required to communicate the information about C over the transmission channel, since in this case the set of features that constitute the vowels is greater in number. Therefore, the higher the perceptual contrast, the higher the feature entropy. The idea is illustrated in Fig. 4. In this figure, F_E exhibited by the

$C_1 = \{/i/, /u/\}$

$N_1 = 2$

$F_1 = \{h, f, b, r, u\}$

F_1	h	f	b	r	u
/i/	1	1	0	0	1
/u/	1	0	1	1	0
p_f/N_1	1	0.5	0.5	0.5	0.5
q_f/N_1	0	0.5	0.5	0.5	0.5

$F_{E_1} = 4$

$C_2 = \{/i/, /\tilde{i}/\}$

$N_2 = 2$

$F_2 = \{h, f, u, n\}$

F_2	h	f	u	n
/i/	1	1	1	0
/ĩ/	1	1	1	1
p_f/N_2	1	1	1	0.5
q_f/N_2	0	0	0	0.5

$F_{E_2} = 1$

Fig. 4. F_E for the two different communities C_1 and C_2. The letters h, f, b, r, u, and n stand for the features "high", "front", "back", "rounded", "unrounded" and "nasalized", respectively.

[f] Note that since feature entropy is a combination of the binary entropy values for all the features, we do not require that $\sum_{f\in F} \frac{p_f}{N} + \sum_{f\in F} \frac{q_f}{N} = 1$.

$C_1 = \{/\tilde{i}/, /\tilde{u}/\}$

$N_1 = 2$

$F_1 = \{h, f, b, r, u, n\}$

F_1	h	f	b	r	u	n
/ĩ/	1	1	0	0	1	1
/ũ/	1	0	1	1	0	1
p_f/N_1	1	0.5	0.5	0.5	0.5	1
q_f/N_1	0	0.5	0.5	0.5	0.5	0

$F_{E_1} = 4$

$C_2 = \{/\tilde{i}/, /u:/\}$

$N_2 = 2$

$F_2 = \{h, f, b, r, u, l, n\}$

F_2	h	f	b	r	u	l	n
/ĩ/	1	0	1	0	1	0	1
/u:/	1	1	0	1	0	1	0
p_f/N_2	1	0.5	0.5	0.5	0.5	0.5	0.5
q_f/N_2	0	0.5	0.5	0.5	0.5	0.5	0.5

$F_{E_2} = 6$

Fig. 5. F_E for the two different communities C_1 and C_2. The letters h, f, b, r, u, l, and n stand for the features "high", "front", "back", "rounded", "unrounded", "long" and "nasalized", respectively.

community C_1 is higher than for the community C_2, since the set of vowels in C_1 is perceptually more distinct than that in C_2.

2.3.1.2. Capturing feature economy

To have more information conveyed using a smaller number of bits, maximization of the combinatorial possibilities of the features used by the constituent vowels in the community C is needed, which is precisely the prediction made by the principle of feature economy. Therefore, the lower the feature entropy, the higher the feature economy. In fact, it is for this reason that in Fig. 5 F_E exhibited by the community C_1 is lower than for the community C_2, since in C_1 the combinatorial possibilities of the features are better utilized by the vowels than in C_2.

2.3.2. *Metric for comparison*

For the purpose of comparison as discussed earlier, we construct a random version of VoNet, namely VoNet$_{\text{rand}}$. Let the frequency of occurrence for each vowel v in UPSID be denoted by f_v. Also, let there be 451 bins, each corresponding to a language in UPSID. f_v bins are then chosen uniformly at random and the vowel v is packed into these bins. Thus, the vowel inventories of the 451 languages corresponding to the bins are generated. In such randomly constructed inventories the effect of none of the forces (perceptual contrast or feature economy) should be prevalent as there is no strict co-occurrence principle that plays a role in the inventory construction. Therefore, these inventories should show a feature entropy no better than is expected by random chance and hence can act as a baseline for all our experiments reported in the following section. The method for constructing the random inventories is summarized in Algorithm 1. $\mathrm{VoNet_{rand}}$ can be built from these randomly generated vowel inventories in a procedure similar to that used for constructing VoNet.

Algorithm 1. Algorithm to construct $\mathrm{VoNet_{rand}}$
begin
  for *each vowel v*
  {
    for $i = 1$ *to* f_v
    {
      Choose one of the 451 bins, corresponding to the languages in UPSID, uniformly at random;
      Pack the vowel v into the bin so chosen if it has not already been packed into this bin earlier;
    }
  }
  Construct $\mathrm{VoNet_{rand}}$, similarly to VoNet, from the new vowel inventories (each bin corresponds to a new inventory);
end

Note that the above construction procedure does not preserve the original vowel inventory sizes, even though it does preserve the occurrence frequency of the vowels. A random inventory construction mechanism, where the inventory sizes are preserved but not the phoneme frequencies, is discussed in Ref. 21 in the context of consonant inventories. Quite unsurprisingly, the characteristics of the random inventories preserving the phoneme frequencies are closer to those of the real inventories than to those generated preserving only the inventory sizes. Similar results are expected in the case of the vowel inventories, i.e. the model which preserves the occurrence frequency of the vowels (Algorithm 1) should be closer to the real inventories than those that just preserve the size of the inventories. This is because the former takes into account at least one of the properties of the vowels (their occurrence frequency) while the latter does not. There can be a third approach to random inventory generation, through *scrambling* (exchanging a random pair of vowels between two inventories subjected to the constraint that no inventory should have the same vowel more than once), where both inventory sizes and vowel frequencies are preserved. Nevertheless, this model is computationally intensive and one can approximate the scrambling process only by allowing a large (though fixed) number of exchanges. The larger the number of exchanges, the better the random model; in other words, the model is sensitive to the number of exchanges. We will show a representative result obtained from this model later in the article [Fig. 11(b)]. However, in the rest of the article, we will use the model introduced in Algorithm 1 as our metric for comparison following the earlier works [21, 22].

3. Experiments and Results

In this section, we describe the experiments performed and the results obtained from the analysis of VoNet. In order to find the co-occurrence patterns in and across the planes of Fig. 1, we define three versions of VoNet, namely $\text{VoNet}_{\text{hub}}$, $\text{VoNet}_{\text{rest}}$ and $\text{VoNet}_{\text{rest}'}$. The construction procedure for each of these versions is presented below.

Construction of $\text{VoNet}_{\text{hub}}$. $\text{VoNet}_{\text{hub}}$ comprises the *hubs*, i.e. the nodes having a very high node weight (frequency).[g] We define a node as a hub if its node weight is greater than 120. Thus $\text{VoNet}_{\text{hub}}$ is a subgraph of VoNet comprising the hubs and the edges interconnecting them. The rest of the nodes (having a node weight less than 120) and edges are removed from the network. We make a choice of this node weight for distinguishing the hubs from the nonhubs by observing the distribution of the occurrence frequency of the vowels illustrated in Fig. 6. The curve shows the frequency of a vowel (y axis) versus the rank of the vowel according to this frequency (x axis) in the log–log scale. The high frequency zone (marked by a circle in the figure) can be easily distinguished from the low frequency one, since there is a distinct gap between the two in the curve.

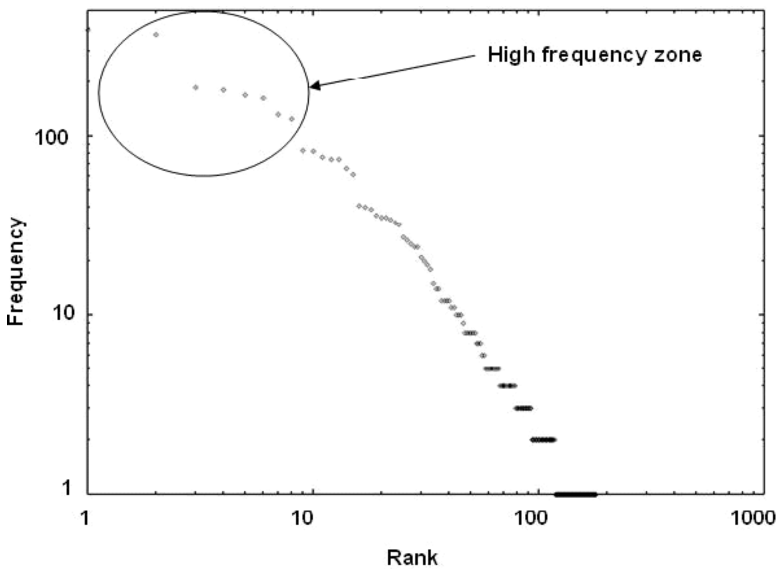


Fig. 6. The frequency (y axis) versus rank (x axis) curve in the log–log scale, illustrating the distribution of the occurrence of the vowels over the language inventories of UPSID.

[g]Hubs are nodes that have a very high degree. In VoNet, nodes having very high frequency also have a very high degree. In particular, all the nodes included in $\text{VoNet}_{\text{hub}}$ have a degree greater than 950 in VoNet, while the average degree of a node in VoNet is only 180 (approx.). Therefore, although the nodes in $\text{VoNet}_{\text{hub}}$ are chosen on the basis of their node weight, we could have chosen them on the basis of the degree and still we would have got the same network.

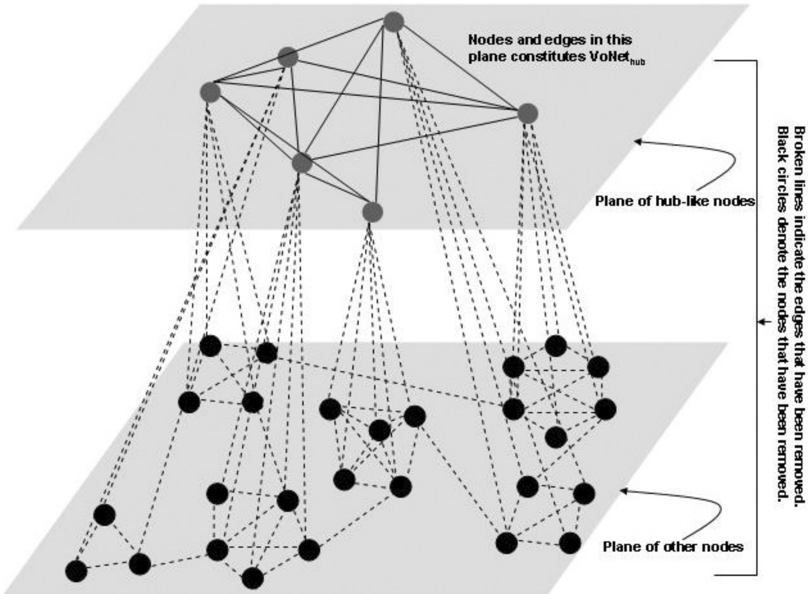


Fig. 7. The construction procedure of VoNet$_{\text{hub}}$ from VoNet.

Figure 7 illustrates how VoNet$_{\text{hub}}$ is constructed from VoNet. Presently, the number of nodes in VoNet$_{\text{hub}}$ is 9 (corresponding to a linguistically relevant set of vowels, which are /i/, /a/, /u/, /ɔ/, /ϵ/, /o/, /e/, /ŏ/ and /ĕ/) and the number of edges is 36.

Construction of VoNet$_{\text{rest}}$. VoNet$_{\text{rest}}$ comprises the same nodes as VoNet. It also has all the edges of VoNet except for those that interconnect the hubs. Figure 8 shows how VoNet$_{\text{rest}}$ can be constructed from VoNet. The number of nodes and of edges in VoNet$_{\text{rest}}$ are 180 and 1293 respectively.[h]

Construction of VoNet$_{\text{rest}'}$. VoNet$_{\text{rest}'}$ also comprises the same nodes as VoNet. It consists of only the edges that connect a hub with a nonhub if the nonhub co-occurs more than 95% of times with the hub. The basic idea behind such a construction is to capture the co-occurrence patterns based on markedness [7] (discussed earlier in the introductory section) that actually defines the cross-planar relationships in Fig. 1. Figure 9 shows how VoNet$_{\text{rest}'}$ can be constructed from VoNet. The number of nodes in VoNet$_{\text{rest}'}$ is 180, while the number of edges is 114.[i] Note that

[h]We have neglected nodes with a node weight less than 3, since these nodes correspond to vowels that occur in less than three languages in UPSID and the communities they form are therefore statistically insignificant. The number 3 has been decided on arbitrarily, based on the manual inspection of the data.

[i]The network does not get disconnected due to this construction since there is always a small fraction of edges that run between the hubs and low-node-weight nonhubs of otherwise disjoint groups.

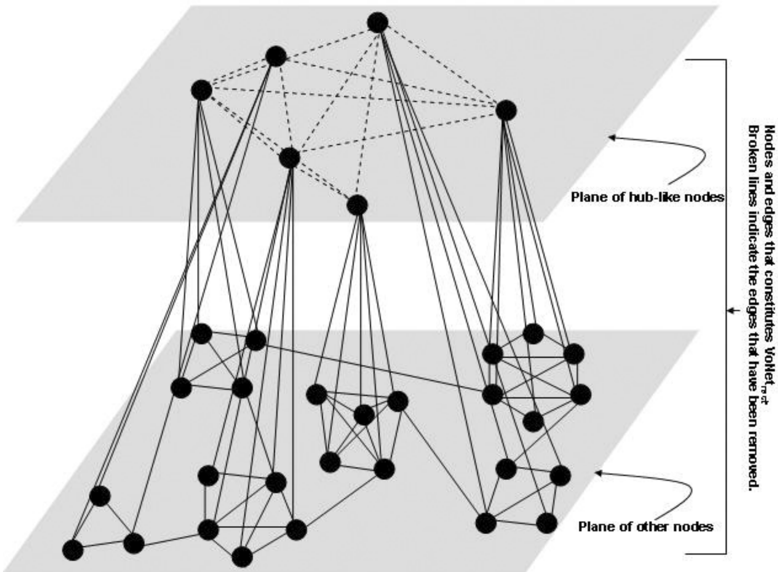


Fig. 8. The construction procedure of VoNet$_{\mathrm{rest}}$ from VoNet.

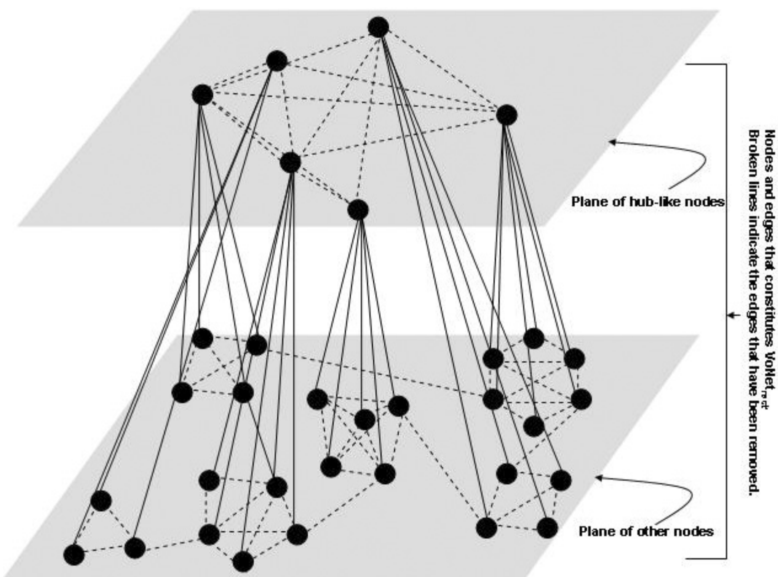


Fig. 9. The construction procedure of VoNet$_{\mathrm{rest}'}$ from VoNet.

since VoNet$_{\mathrm{rest}'}$ has edges running between the frequent (i.e. unmarked) and the infrequent (i.e. marked) vowels, a community structure analysis of this network is expected to reveal a relationship between marked and unmarked pairs of vowels that co-occur frequently.

Table 1. Vowel communities obtained from clustering of $\mathrm{VoNet}_{\mathrm{hub}}$. The contrasting features separated by slashes are shown within parentheses. Comma-separated entries represent the features that are in use from the three respective classes, namely the height, the backness and the roundedness.

Community	Features in contrast
/i/, /a/, /u/	(low/high), (front/central/back), (unrounded/rounded)
/e/, /o/	(higher-mid/mid), (front/back), (unrounded/rounded)

Table 2. Some of the vowel communities obtained from $\mathrm{VoNet}_{\mathrm{rest}}$.

Community	Features in Common
/ĩ/, /ã/, /ũ/	nasalized
/ĩ:/, /ã:/, /ũ:/	long, nasalized
/i:/, /u:/, /a:/, /o:/, /e:/	long

Table 3. Some of the vowel communities obtained from $\mathrm{VoNet}_{\mathrm{rest}'}$. Comma-separated entries represent the features that are in use from the three respective classes, namely the height, the backness and the roundedness.

Community	Features in common
/i/, /ĩ/	high, front, unrounded
/a/, /ã/	low, central, unrounded
/u/, /ũ/	high, back, rounded

We separately apply the MRad algorithm to $\mathrm{VoNet}_{\mathrm{hub}}$, $\mathrm{VoNet}_{\mathrm{rest}}$ and $\mathrm{VoNet}_{\mathrm{rest}'}$ in order to obtain the respective vowel communities. Representative communities from $\mathrm{Vonet}_{\mathrm{hub}}$, $\mathrm{VoNet}_{\mathrm{rest}}$ and $\mathrm{VoNet}_{\mathrm{rest}'}$ are noted in Tables 1, 2 and 3 respectively.

Tables 1, 2 and 3 indicate that the communities in $\mathrm{VoNet}_{\mathrm{hub}}$ are formed based on the principle of perceptual contrast, whereas the formation of the communities in $\mathrm{VoNet}_{\mathrm{rest}}$ and $\mathrm{VoNet}_{\mathrm{rest}'}$ is largely governed by feature economy. In the rest of this section, we focus mainly on verifying the above argument. For this reason we present a detailed study of the co-occurrence principles of the communities obtained from $\mathrm{VoNet}_{\mathrm{hub}}$, $\mathrm{VoNet}_{\mathrm{rest}}$ and $\mathrm{VoNet}_{\mathrm{rest}'}$. In each case we compare the results with those of $\mathrm{VoNet}_{\mathrm{rand}}$ obtained from Algorithm 1. Note that the threshold η of the MRad algorithm can take different values and each such value results in a set of communities. In all the experiments that we report henceforth, the value of η has been varied from 1.8 to 0.005 (each time the value is reduced by a factor of $\frac{1}{1.2}$) in order to obtain different sets of communities.

3.1. *Co-occurrence principles of the communities of $\mathit{VoNet}_{\mathbf{hub}}$*

The random counterpart of $\mathrm{VoNet}_{\mathrm{hub}}$ (henceforth $\mathrm{VoNet}_{\mathrm{Rhub}}$) is constructed from $\mathrm{VoNet}_{\mathrm{rand}}$ following the steps that were used to construct $\mathrm{VoNet}_{\mathrm{hub}}$ from VoNet.

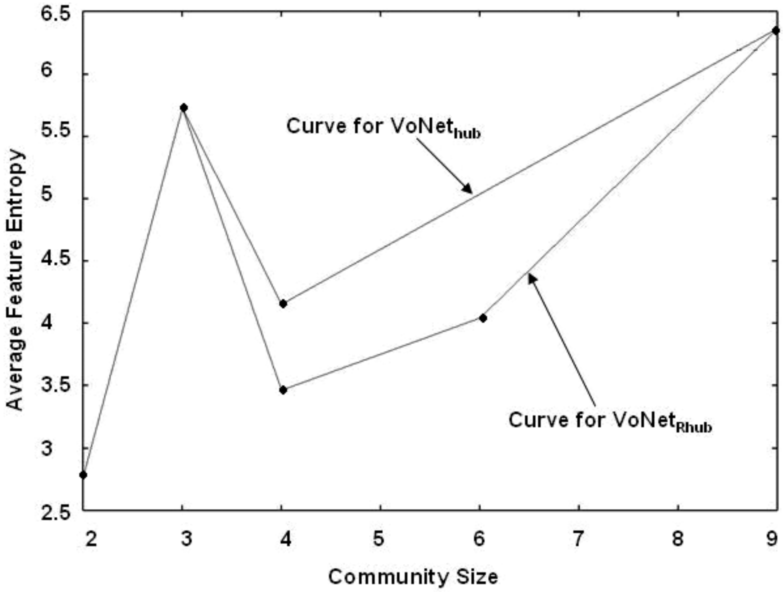


Fig. 10. Curves showing the average feature entropy of the communities of a particular size versus the community size for $\text{VoNet}_{\text{hub}}$ as well as $\text{VoNet}_{\text{Rhub}}$.

Figure 10 illustrates, for all the communities obtained from the clustering of $\text{VoNet}_{\text{hub}}$ and $\text{VoNet}_{\text{Rhub}}$, the average feature entropy exhibited by the communities of a particular size[j] (y axis) versus the community size (x axis).

A closer inspection of Fig. 10 immediately reveals that the feature entropy exhibited by the communities of $\text{VoNet}_{\text{hub}}$ is higher as compared to the random version of the same. The two curves finally intersect because eventually, for a low value of η, all the nodes in $\text{VoNet}_{\text{hub}}$ and $\text{VoNet}_{\text{Rhub}}$ form a single connected component, i.e. a single cluster. Since the set of hubs, which are defined solely in terms of node weight, is identical for VoNet and $\text{VoNet}_{\text{rand}}$, the feature entropy of the cluster formed by all the hubs is also identical in both the cases.

Nevertheless, the data points that appear on these curves are fairly less in number and hence Fig. 10 alone is not sufficient to establish that the communities in $\text{VoNet}_{\text{hub}}$ are formed based on the principle of perceptual contrast. Another possible way to investigate the problem would be to look into the co-occurrence principles of the smaller vowel inventories (of size ≤ 4), since they mostly comprise the hubs. Table 4, for instance, shows the number of occurrences of the members of the community formed by /i/, /a/ and /u/, as compared to the average occurrence of other vowels, in the inventories of sizes 3 and 4. The figures in the table point to

[j]Let there be n communities of a particular size k picked up at all the different values of η. The average feature entropy of the communities of size k is therefore $\frac{1}{n}\sum_{i=1}^{n} F_{E_i}$, where F_{E_i} signifies the feature entropy of the ith community.

Table 4. Frequency of occurrence of the members of the community /i/, /a/ and /u/, as compared to the frequency occurrence of other vowels, in smaller inventories. The last column indicates the average number of times that a vowel other than /i/, /a/ and /u/ occurs in the inventories of sizes 3 and 4.

Inventory size	Number of inventories	Occurrence of /i/	Occurrence of /a/	Occurrence of /u/	Average occurrence of other vowels
3	23	15	21	12	3
4	25	19	24	11	3

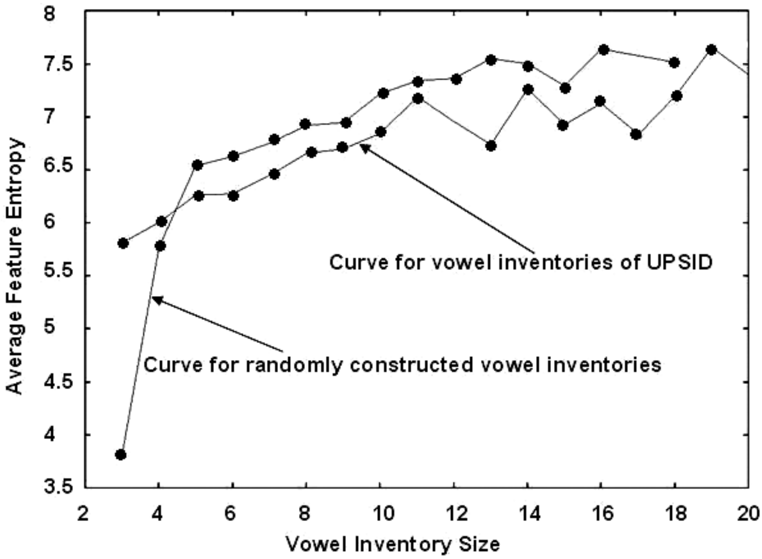
the fact that the smaller inventories can be assumed to be good representatives of the communities obtained from $\mathrm{VoNet}_{\mathrm{hub}}$. We therefore compare the average feature entropy of these inventories as a whole with their random counterparts (obtained from Algorithm 1). Figure 11(a) illustrates the result of this comparison. The curves depict the average feature entropy of the vowel inventories of a particular size (y axis) versus the inventory size (x axis). The two different plots compare the average feature entropy of the inventories obtained from UPSID with that of the randomly constructed ones. The figure clearly shows that the average feature entropy of the vowel inventories of UPSID is substantially higher for inventory sizes 3 and 4 than that of those constructed randomly.

Figure 11(b) presents the results of experiments similar to those above except for the fact that in this case we employ the process of scrambling to construct the random inventories. For 0.1 million exchanges, the average feature entropy curve for the random inventories is quite close to that of the real ones; as we increase the number of exchanges to 1 million, the two curves move far apart from each other, indicating that the randomness increases with the increase in the number of pairs of vowels being scrambled. Note that in this case also the feature entropy exhibited by the smaller inventories is (most of the time) higher than that of those generated randomly. However, the results described here should be interpreted with caution because of the approximation involved in the model.

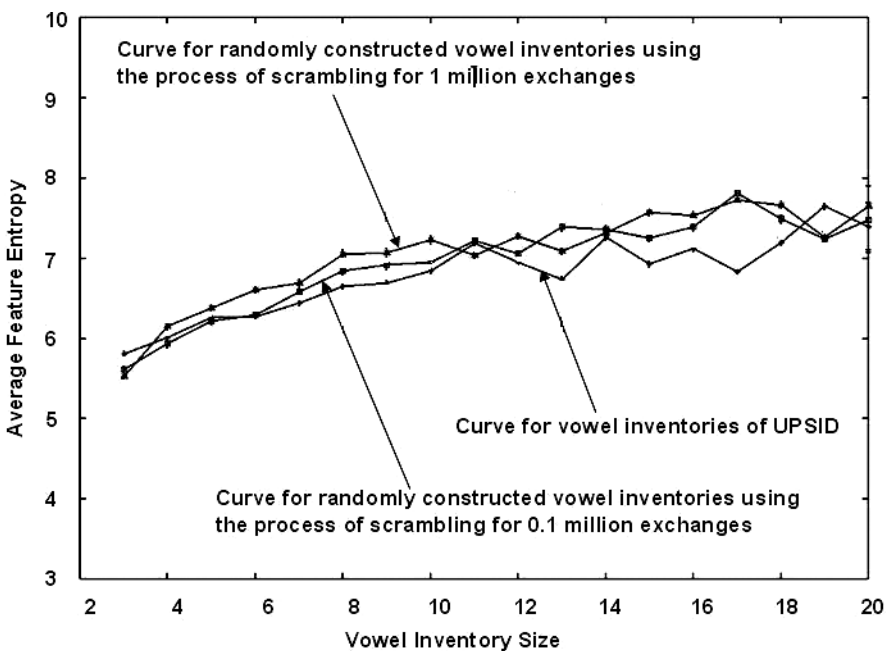
The results presented in Figs. 10 and 11 together confirm that the communities in $\mathrm{VoNet}_{\mathrm{hub}}$ are formed based on the principle of maximal perceptual contrast.

3.2. *Co-occurrence principles of the communities of $\mathit{VoNet}_{\mathbf{rest}}$*

In this subsection, we investigate whether or not the communities obtained from $\mathrm{VoNet}_{\mathrm{rest}}$ are better in terms of feature entropy than they would have been if the vowel inventories had evolved just by chance. We construct the random version of $\mathrm{VoNet}_{\mathrm{rest}}$ (henceforth $\mathrm{VoNet}_{\mathrm{Rrest}}$) from $\mathrm{VoNet}_{\mathrm{rand}}$ and apply the MRad algorithm to it so as to obtain the communities. Figure 12 illustrates, for all the communities obtained from the clustering of $\mathrm{VoNet}_{\mathrm{rest}}$ and $\mathrm{VoNet}_{\mathrm{Rrest}}$, the average feature entropy exhibited by the communities of a particular size (y axis) versus the community size (x axis). The curves in the figure make it quite clear that the average feature entropy exhibited by the communities of $\mathrm{VoNet}_{\mathrm{rest}}$ is substantially lower



(a)



(b)

Fig. 11. Curves showing the average feature entropy of the vowel inventories of a particular size versus the inventory size. (a) The two different plots compare the average feature entropy of the inventories obtained from UPSID with that of the randomly constructed ones (obtained from Algorithm 1). (b) The different plots compare the average feature entropy of the inventories obtained from UPSID with that of the randomly constructed ones through the method of scrambling.

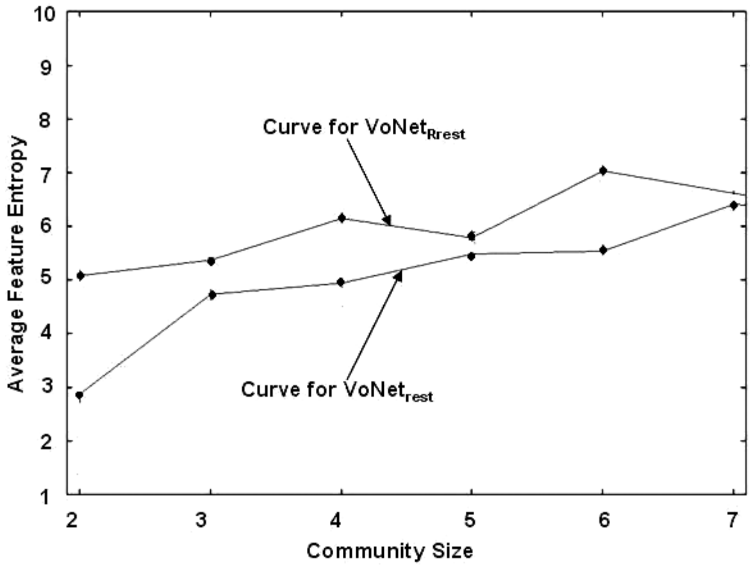


Fig. 12. Curves showing the average feature entropy of the communities of a particular size versus the community size for $\text{VoNet}_{\text{rest}}$ as well as $\text{VoNet}_{\text{Rrest}}$.

than that of $\text{VoNet}_{\text{Rrest}}$ (especially for a community size ≤ 7). As the community size increases, the difference between the average feature entropy of the communities of $\text{VoNet}_{\text{rest}}$ and that of the communities of $\text{VoNet}_{\text{Rrest}}$ gradually diminishes. This is mainly because of the formation of a giant community, which is similar for both $\text{VoNet}_{\text{rest}}$ and $\text{VoNet}_{\text{Rrest}}$.

The above result indicates that the driving force behind the formation of the communities of $\text{VoNet}_{\text{rest}}$ is the principle of feature economy. It is important to mention here that the larger vowel inventories, which usually comprise the communities of $\text{VoNet}_{\text{rest}}$, also exhibit feature economy to a large extent. This is reflected through Fig. 11(a), where all the real inventories of size ≥ 5 have a substantially lower average feature entropy than the randomly generated ones.

3.3. *Co-occurrence principles of the communities of* $\textit{VoNet}_{\textbf{rest}'}$

In this subsection, we compare the feature entropy of the communities obtained from $\text{VoNet}_{\text{rest}'}$ with that of its random counterpart, $\text{VoNet}_{\text{Rrest}'}$ (constructed from $\text{VoNet}_{\text{rand}}$). Figure 13 shows the average feature entropy exhibited by the communities of a particular size (y axis) versus the community size (x axis) for both $\text{VoNet}_{\text{rest}'}$ and $\text{VoNet}_{\text{Rrest}'}$. The curves in the figure make it quite clear that the average feature entropy exhibited by the communities of $\text{VoNet}_{\text{rest}'}$ is substantially lower than that of $\text{VoNet}_{\text{Rrest}'}$. This result immediately reveals that it is again feature economy that plays a key role in the emergence of the communities of $\text{VoNet}_{\text{rest}'}$.

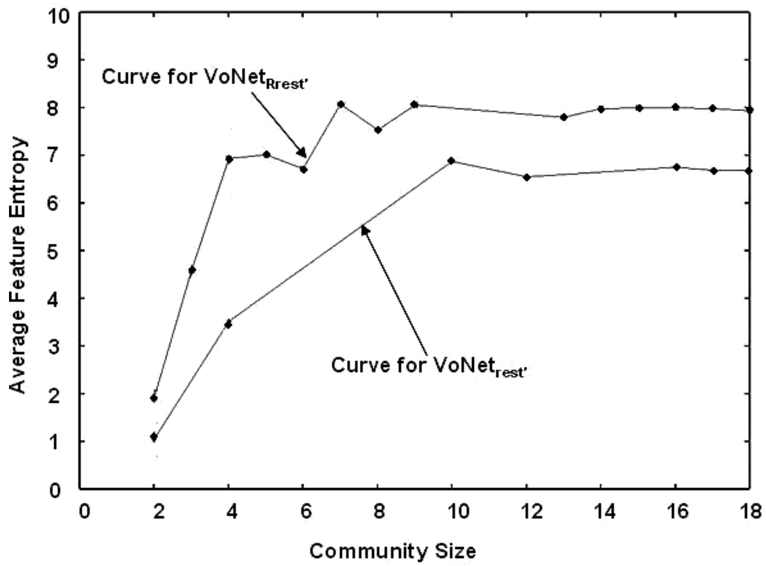


Fig. 13. Curves showing the average feature entropy of the communities of a particular size versus the community size for $\mathrm{VoNet}_{\mathrm{rest}'}$ as well as $\mathrm{VoNet}_{\mathrm{Rrest}'}$.

4. Conclusion

In this paper, we have explored the co-occurrence principles of vowels, across the inventories of the world's languages. We started with a concise review of the available literature on vowel inventories. We proposed an automatic procedure for capturing the co-occurrence patterns of the vowels across languages. We also discussed the notion of feature entropy, which immediately allows us to validate the explanations of the organizational principles of the vowel inventories furnished by the earlier researchers.

Some of our important findings from this work are:

- The smaller vowel inventories (corresponding to the communities of $\mathrm{VoNet}_{\mathrm{hub}}$) tend to be organized based on the principle of maximal perceptual contrast;
- On the other hand, the larger vowel inventories (mainly comprising the communities of $\mathrm{VoNet}_{\mathrm{rest}}$) reflect a considerable extent of feature economy;
- Co-occurrences based on markedness are prevalent across vowel inventories (captured through the communities of $\mathrm{VoNet}_{\mathrm{rest}'}$) and their emergence is again a consequence of feature economy.

It may be noted that the results presented are quite robust and can be generalized since the database (UPSID) that we have used for our experiments constitutes a genetically balanced sample of the world's languages. We selected UPSID mainly for two reasons: (i) it is the largest database of this type that is currently available; (ii) it has been constructed by selecting one language each from moderately distant language families, which ensures a considerable degree of genetic balance. Therefore, as the data has almost a negligible bias toward any particular family, we can expect that the conclusions regarding the organizational principles of the vowel inventories that we draw here can be generalized over a large majority of the languages of the world. The same reason also makes the conclusions quite insensitive to the languages that are considered as long as the genetic balance is maintained.

The principles of feature economy and maximal perceptual contrast operate at a certain level of cognition, where speech sounds are assumed to be encoded in terms of relatively abstract elements (features) within a linguistic system. While feature economy tends to organize the linguistic data into a small number of groups, perceptual contrast tends to increase this number so as to minimize the level of confusion. If U is the set of linguistic units, and C the categories that characterize these units, then feature economy may be expressed as "maximize U/C" (as suggested in Ref. 6) and perceptual contrast as "minimize U/C." It is the interplay of these two optimization principles that shapes the structure of the vowel inventories. While the smaller inventories can afford to choose perceptually distinct vowels (since there are very few vowels to be learnt and hence the effort of learnability is low), the larger inventories tend to be economical so that the effort of learnability does not increase considerably (because in this case, though there are many vowels to be learnt, the number of features that are actually to be learnt is smaller).

In the preceding sections, we have mainly emphasized the analysis of the co-occurrence principles of the vowel inventories of the world's languages. An issue that draws attention is how the forces of perceptual contrast and feature economy have interacted to cause the emergence of the human vowel systems. One possible way of answering this question is to have a growth model for the network, where the growth takes place owing to the optimization of a function (see for example Ref. 4), which involves the above forces and also accounts for the observed regularities displayed by the vowel inventories. It is worthwhile to mention here that though most of the mechanisms of network growth [24] rely on preferential attachment-based rules [1], there are scenarios which suggest that additional optimizing constraints need to be imposed on the evolving network so as to match its emergent properties with empirical data [28, 29]. Such a growth model based on some optimization technique can then shed enough light on the real dynamics that went on in the evolution of the vowel inventories. We look forward to developing the same as a part of our future work.

Acknowledgments

A. M. and M. C. would like to thank Microsoft Research India and Media Lab Asia respectively for financial assistance. All the authors would like to extend their gratitude to the anonymous referees for their valuable comments and suggestions.

References

[1] Barabási, A.-L. and Albert, R., Emergence of scaling in random networks, *Science* **286** (1999) 509-512.

[2] de Boer, B., Self-organisation in vowel systems, *J. Phonetics* **28** (2000) 441–465.

[3] Boersma, P., Functional phonology, Doctoral thesis, University of Amsterdam (Holland Academic Graphics, The Hague, 1998).

[4] Ferrer i Cancho, R. and Solé, R. V., Optimization in complex networks, in *Lecture Notes in Physics* (Springer, Berlin, 2003), pp. 114–125.

[5] Choudhury, M., Mukherjee, A., Basu, A. and Ganguly, N., Analysis and synthesis of the distribution of consonants over languages: A complex network approach, in *Proc. COLING–ACL* (Sydney, 2006), pp. 128–135.

[6] Clements, G. N., Feature economy in sound systems, *Phonology* **20** (2003) 287–333.

[7] Clements, G. N., The role of features in speech sound inventories, in E. Raimy and C. Cairns (eds.), *Contemporary Views on Architecture and Representations in Phonological Theory* (MIT Press, 2008).

[8] de Groot, A. W., Phonologie und Phonetik als funktionswissenschaften, *Travaux du Cercle Linguistique de Prague* **4** (1931) 116–147.

[9] Hinskens, F. and Weijer, J., Patterns of segmental modification in consonant inventories: a cross-linguistic study, *Linguistics* **41** (2003) 1041–1084.

[10] Jakobson, R., *Kindersprache, aphasie und allgemeine lautgesetze* (Uppsala, 1941), reprinted in *Selected Writings of I. Mouton* (The Hague, 1962), pp. 328–401.

[11] Ke, J., Ogura, M. and Wang, W. S.-Y., Optimization models of sound systems using genetic algorithms, *Comput. Linguistics* **29** (2003) 1–18.

[12] Korn, G. A. and Korn, T. M., *Mathematical Handbook for Scientists and Engineers: Definitions, Theorems, and Formulas for Reference and Review* (Dover, New York, 2000), pp. 613–614.

[13] Ladefoged, P. and Maddieson, I., *Sounds of the World's Languages* (Blackwell, Oxford, 1996).

[14] Liljencrants, J. and Lindblom, B., Numerical simulation of vowel quality systems: The role of perceptual contrast, *Language* **48** (1972) 839–862.

[15] Lindblom, B., Phonetic universals in vowel systems, *Exp. Phonol.* (1986) 13–44.

[16] Lindblom, B. and Maddieson, I., Phonetic universals in consonant systems, in Hyman, L. M. and Li, C. N. (eds.), *Language, Speech, and Mind* (Routledge, London, 1988), pp. 62–78.

[17] Maddieson, I., *Patterns of Sounds* (Cambridge University Press, 1984).

[18] Martinet, A., *Ѐconomie des changements phonétiques* (A. Francke, Berne, 1955).

[19] Broe, M., A generalized information-theoretic measure for systems of phonological classification and recognition, in *Proc. Second Meeting of the ACL Special Interest Group in Computational Phonology* (1996).

[20] Miller, E., Matsakis, N. and Viola, P., Learning from one example through shared densities on transforms, in *Proc. IEEE Conference on Computer Vision and Pattern Recognition* (2000), pp. 464–471.

[21] Mukherjee, A., Choudhury, M., Basu, A. and Ganguly, N., Modeling the co-occurrence principles of the consonant inventories: A complex network approach, *Int. J. Mod. Phys. C* **18** (2007) 281–295.

[22] Mukherjee, A., Choudhury, M., Basu, A. and Ganguly, N., Redundancy ratio: An invariant property of the consonant inventories of the world's languages, in *Proc. 45th Annual Meeting of the Association of Computational Linguistics* (2007), pp. 104–111.

[23] Mukherjee, A., Choudhury, M., Basu, A. and Ganguly, N., Self-organization of the sound inventories: Analysis and synthesis of the occurrence and co-occurrence networks of consonants, *J. Quant. Ling.*, accepted [arXiv.org/physics/0606132].

[24] Newman, M. E. J., The structure and function of complex networks, *SIAM Rev.* **45** (2003) 167–256.
[25] Radicchi, F., Castellano, C., Cecconi, F., Loreto, V. and Parisi, D., Defining and identifying communities in networks, *PNAS* **101** (2003) 2658–2663.
[26] Schwartz, J.-L., Boë, L.-J., Vallée, N. and Abry, C., The dispersion-focalization theory of vowel systems, *J. Phonetics* **25** (1997) 255–286.
[27] Shannon, C. E., and Weaver, W., *The Mathematical Theory of Information* (University of Illinois Press, Urbana, 1949).
[28] Solé, R. V., Pastor-Satorras, R., Smith, E. and Kepler, T., A model of large-scale proteome evolution, *Adv. Complex Syst.* **5** (2002) 43–54.
[29] Vázquez, A., Flammini, A., Maritan, A. and Vespignani, A., Modeling of protein interaction networks, *Complexus* **1** (2003) 38–44.
[30] Wang, W. S.-Y., The basis of speech. Project on linguistic analysis reports (University of California, Berkeley, 1968), reprinted in *The Learning of Language* (1971).
[31] Warfield, S. K., Rexilius, J., Huppi, P. S., Inder, T. E., Miller, E. G., Wells, W. M., Zientara, G. P., Jolesz, F. A. and Kikinis, R., A binary entropy measure to assess nonrigid registration algorithm, in *Proc. Fourth Int. Conf. Medical Image Computing and Computer Assisted Intervention* (2001), pp. 266–274.