Optimization Models of Sound Systems Using Genetic Algorithms

Jinyun Ke*
City University of Hong Kong

Mieko Ogura†
Tsurumi University
University of California at Berkeley

William S.-Y. Wang‡
City University of Hong Kong
University of California at Berkeley

In this study, optimization models using genetic algorithms (GAs) are proposed to study the configuration of vowels and tone systems. As in previous explanatory models that have been used to study vowel systems, certain criteria, which are assumed to be the principles governing the structure of sound systems, are used to predict optimal vowels and tone systems. In most of the earlier studies only one criterion has been considered. When two criteria are considered, they are often combined into one scalar function. The GA model proposed for the study of tone systems uses a Pareto ranking method that is highly applicable for dealing with optimization problems having multiple criteria. For optimization of tone systems, perceptual contrast and markedness complexity are considered simultaneously. Although the consistency between the predicted systems and the observed systems is not as significant as those obtained for vowel systems, further investigation along this line is promising.

1. Introduction

Studies of the universal characteristics of sound systems in human languages can be pursued according to two different approaches, an inductive approach and a deductive one. The inductive approach involves analyzing the database built from a survey of a large number of languages to arrive at a list of "universal" features that can be widely observed in the database. The deductive approach hypothesizes a number of principles related to speech production and perception processes and predicts possible systems using these principles. These two approaches, however, are often interwoven. The principles hypothesized by the deductive approach are modified or falsified by comparing the predictions with the results from the inductive analysis of real language systems. At the same time, the ultimate aim for inductive analysis is to seek intrinsic mechanisms and principles of human speech to explain the universals found in real systems.

For the inductive approach in phonological studies, there are two large-scale databases available. One is the Stanford Phonology Archiving (SPA) Project (Vihman 1977), which initially included 196 languages and was extended to 209 languages in

* Department of Electronic Engineering, City University of Hong Kong, Hong Kong. E-mail: jyke@ee.cityu.edu.hk
† Linguistics Laboratory, Tsurumi University, Yokohama, Japan; Project on Linguistic Analysis, University of California at Berkeley. E-mail: ogura-m@tsurumi-u.ac.jp
‡ Department of Electronic Engineering, City University of Hong Kong, Hong Kong; Project on Linguistic Analysis, University of California at Berkeley. E-mail: eewsyw@uxmail.cityu.edu.hk

1978. The other is the Phonological Segment Inventory Database (UPSID) (Maddieson 1984) at the University of California at Los Angeles, which initially included 371 languages and was later extended to 451 languages (Maddieson and Precoda 1990; Ladefoged and Maddieson 1996). Many typological studies have been carried out based on these two databases. For example, in studying vowel systems, Crothers (1978) reported an analysis using the SPA database. Ladefoged and Maddieson (1990) and Schwartz et al. (1997b) reported comprehensive analyses for the vowels systems in UPSID.

Along with typological studies of the languages in these databases, explanatory models, which attempt to explore the intrinsic reasons for structures and universals, have also been proposed. In the study of vowel systems, the principle of **maximal perceptual contrast** has a long tradition in linguistics (Jakobson 1941; Wang 1968). This principle suggests that a vowel system tends to achieve a maximum contrast among the vowels in the system. A number of numerical studies adopting this principle have been proposed (Liljencrants and Lindblom 1972; Crothers 1978; Lindblom 1986). Lindblom (1986) proposed the **sufficient perceptual contrast** principle, under which more systems are predicted to be consistent with natural systems than is predicted by the maximal perceptual contrast principle. Boë, Schwartz, and Vallée (1994) and Schwartz et al. (1997a) added a new consideration called the **focalization** principle that is based on the observation that vowels with strong formant convergence would be perceptually preferred. More recently, de Boer (1997, 2000, 2001) proposed a synthesized model in which agents interact with each other through iterative imitation games. With explicit optimization, agents can develop coherent vowel systems that are close to real systems.

All of the works cited above are concerned with vowel systems. Far fewer studies are reported on other components of a sound system, including consonants, tones (in tone languages), and pitch accent (in non–tone languages), than on vowels. Lindblom and Maddieson (1988) reported a study on phonetic universals in consonant systems using data from UPSID. They proposed that the structure of consonant systems does not arise from a single principle such as the maximization of perceptual contrast. Instead, articulatory factors interact with perceptual factors. According to their proposal, consonant inventories tend to evolve so as to achieve maximal perceptual distinctiveness at minimum articulatory cost.

There are some inductive studies on the universals of tone systems as well. For example, Maddieson (1978) reviewed the phonological universals of tones by analyzing data from SPA. Also, Cheng (1973) reported a detailed analysis of the tone systems in Chinese dialects. We have not, however, found any explanatory models that apply a deductive approach for tone systems in the way that such an approach has been applied for vowel systems.

More recently, Redford, Chen, and Miikkulainen (2001) reported their studies on the universal and variations of syllable structures (i.e., the combinations of vowels and consonants). They developed a computational model based on a version of the genetic algorithm (GA) (Holland 1975) to simulate the emergence of syllable systems in a language. A set of functional constraints related to perceptual distinctiveness and articulatory ease are taken into account as optimization objectives.

In this study, we report some optimization models using GAs to study optimal vowel and tone systems. In these models, the optimal systems are derived from the models based on various explicit optimization criteria and compared with observed systems. First, in the study of vowel systems, we compare two sets of criteria, one considering only the principle of maximal perceptual contrast (Liljencrants and Lindblom 1972), and the other considering both the intervowel perceptual distance and the intravowel spectral salience, that is, the dispersion-focalization principle proposed by Schwartz et al. (1997a). In the second set of criteria, the two objectives, that is, inter-

vowel perceptual distance and the intravowel spectral salience, are combined into a scalar function. In comparing our results with those of earlier studies, we find that the GA models demonstrate the effectiveness of the GA method in identifying the optimal vowel systems based on the above criteria.

Second, we apply the GA method to study tone systems. In our application, two objectives (maximum perceptual contrast and minimum markedness complexity) are taken into account to predict the "optimal" tone systems. Instead of combining the two objectives into one fitness function, we use a multi-objective GA (MOGA) model in which a Pareto ranking method is applied for the fitness function. For comparison, we also try a simple GA model that uses only perceptual distance as the optimization criterion. The predicted systems are compared with the real systems for the two sets of criteria.

In the following parts of the article, Section 2 gives a brief introduction to a simple GA and a MOGA. Section 3 reports the simulation we performed for vowel systems and comparisons with previous reports. Section 4 introduces our models for tone systems, together with a new analysis of an available tone systems database. Conclusions and discussion are given in Section 5.

2. Introduction to Genetic Algorithms

2.1 Simple Genetic Algorithm

GAs were first proposed by John Holland in the 1960s (Holland 1975) and have become widely used in various disciplines. The original goal of Holland's GAs was to study the phenomena of adaptation formally by importing the mechanisms of natural adaptation into computer simulation models. Most of the current applications of GAs, however, are used for specific optimization problems in which the focus is on the derivation of optimal solutions to the problem rather than the process of adaptation.

The basic idea of GAs is based on "natural selection," the principle of "survival of the fittest," which assumes that the individual that is better fitted to a particular environment produces more offspring than others in that environment that are less well suited for it; its "fit" genes are then transmitted to the next generation. A GA operates on a population of chromosomes, each generating a potential solution to the studied problem. The process of a traditional simple GA is as follows: At the beginning of the algorithm, a population is randomly initialized, and the fitness of each chromosome is evaluated according to an objective function (also called a fitness function). A number of chromosomes are selected as parents from the population according to their fitness, and parents then undergo crossover and mutation to produce offspring with certain probabilities. Offspring with better fitness are then inserted into the population, replacing the inferior chromosomes in the previous generation. With this replacement, usually the population size is kept constant. This cycle is repeated for a given number of generations, or stopped when a solution obtained is deemed optimal. This process leads to the evolution of a population in which the individuals are more and more suited to their environment, just as in natural adaptation. Because of its global search mechanism, a GA model usually can find global optimal solutions in a more efficient way than traditional optimization methods.

2.2 Multi-objective Genetic Algorithm

In a traditional GA, the fitness function deals only with one optimization objective. Many practical problems, however, are concerned with several equally important (and usually conflicting) objectives. These types of problems are called multi-objective or multicriteria optimization problems (MOPs) (Stadler 1988).

Human language is an instance of an MOP. A language system is constrained by many demands and requirements. We can consider the current language system to be the product of an optimization process based on such constraints. The constraints can be divided mainly into three categories—the speaker constraint, the listener constraint, and the learner constraint—which often lead to different directions of development for the system. For example, for a sound system, the requirements that arise from speaking and listening often conflict with each other. A sound that is easy for a speaker to produce may not be easy for the listener to perceive. Similarly, perceptually distinctive sounds may be difficult to pronounce. A system with a high perceptual contrast may have a high production cost at the same time, as is the case with the consonant set [ɑɭ˙ k' ts ʈ m r ɻ] suggested by Ohala in questioning the effectiveness of the principle of maximum perceptual difference in explaining consonant universals (Lindblom and Maddieson 1988).

We can see the effects of such a tug-of-war between conflicting requirements in various aspects of a language system. For example, in the perception of tones, a completely level tone is the easiest, from a psychophysical viewpoint, to differentiate from nonlevel tones. It requires much effort on the part of a speaker, however, to produce a perfectly level tone. As a consequence of accommodating a speaker's effort, the listeners will shift their linguistic perception boundary between level and rising tones away from the psychophysical boundary, to allow the speaker some freedom in articulation (Wang 1976).

Also, in syntax, a language with free word order may give the speaker a high degree of flexibility in constructing sentences; however, it places the burden on the listener to figure out the relationships among the words. This is solved by signaling the roles of words by various case markers. If the case marking system is too complex, however, it will be hard for children to learn as they acquire the language. Therefore there may exist a balance point among the three different constraints involved.

The most distinctive characteristic of an MOP is that it does not have one singular optimal solution, but rather a set of nondominated,[1] alternative solutions, which is often called the Pareto-optimal set. Recently a set of algorithms, called multi-objective genetic algorithms, have been developed specifically to solve such multi-objective problems. MOGAs have received much attention, and many scientific and engineering applications employing them have been reported (Fonseca and Fleming 1998; Van Veldhuizen and Lamont 2000).

The simplest and most common way to tackle an MOP is to combine its several objectives into one scalar function as the fitness function. Different objectives in the problem are given different weights based on some a priori knowledge (Stadler 1988). (Early studies on sound system optimization with multiple criteria, such as Redford, Chen, and Miikkulainen [2001] and Schwartz et al. [1997a], adopted this approach.) Such knowledge is very often unavailable, however, and most of the time the weights are chosen by trial and error. Thus the performance of the algorithm usually is sensitive to or biased by the weights assigned to the objectives.

Within the GA approach, another method called Pareto ranking is often used in the fitness evaluation. The several objective values of a particular chromosome are maintained as a vector, instead of being combined by means of a scalar function into one single fitness value. The fitness of a chromosome is determined by its ranking

1 Assuming a minimization problem with p objectives, dominance is defined as follows: x_1 is said to dominate x_2 (or x_2 to be inferior to x_1), if the fitness of x_1, $f(x_1)$, is partially less than the fitness of x_2, $f(x_2)$, that is, $f_i(x_1) \leq f_i(x_2)$, $\forall i \in \{1, 2, \ldots, p\}$; and $f_i(x_1) < f_i(x_2)$, $\exists i \in \{1, 2, \ldots, p\}$. A nondominated solution is a solution such that there are no other solutions whose objectives are all better than its.

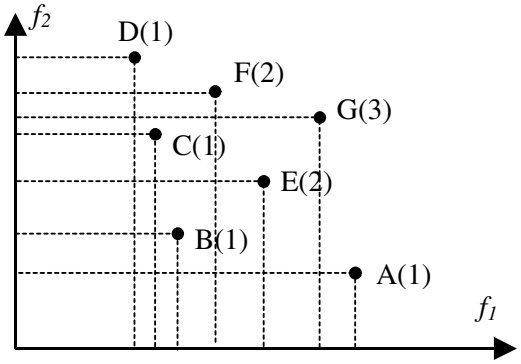

Figure 1
An illustration of Goldberg's Pareto ranking method. The numbers in parentheses represent the ranks for the chromosomes.

in the population, which is obtained from comparing its objective vector with that of others. There are many methods for ranking the objective vectors of chromosomes, and in this study we use the one proposed in Goldberg (1989). Goldberg's method assigns ranks according to the following procedures. First, the nondominated chromosomes from the whole population are found, rank 1 is assigned to them, and they are removed from further consideration in the ranking process. Then, another set of chromosomes is found from the remaining population that is now nondominated (because of the removal of those assigned rank 1) and is assigned rank 2, and so forth. To illustrate the algorithm, Figure 1 gives an example. Points A, B, C, D, E, F, and G represent candidate solutions to a problem whose goal is to minimize two objectives. The objective values of the solutions are shown in the f_1-f_2 plane. According to Goldberg's method, solutions A, B, C, and D are all nondominated and therefore assigned rank 1 and removed. Solutions E and F are now nondominated, so both have rank 2; G has rank 3, the worst rank.

3. Optimization Model for Vowel Systems

In this study, we use a simple GA model to search for the optimal configuration of systems of simple vowels. We apply various optimization criteria proposed by earlier studies (Liljencrants and Lindblom 1972; Lindblom 1986; Schwartz et al. 1997a) in our GA model and compare the predictions produced under the various criteria. We also use two sets of vowel inventories to provide another type of comparisons. The predictions are also compared with observed systems.

3.1 Implementation of the GA Model

The GA model consists of a population with a number of chromosomes, each representing a possible vowel system. Each vowel is encoded according to the three primary parameters of vowel articulation, that is, tongue height, tongue backness, and lip roundness. The first two articulatory parameters are supposed to be continuous within the range of [0,1], whereas the last parameter is a binary value. Though this encoding method allows an infinite number of vowels, following previous studies (Lindblom 1986; Schwartz et al. 1997a), we assume there is a limited inventory of prototypes from which the system can select candidate vowels. Only normal plain vowels are considered. The encoding of the prototypical vowels is designed according to the

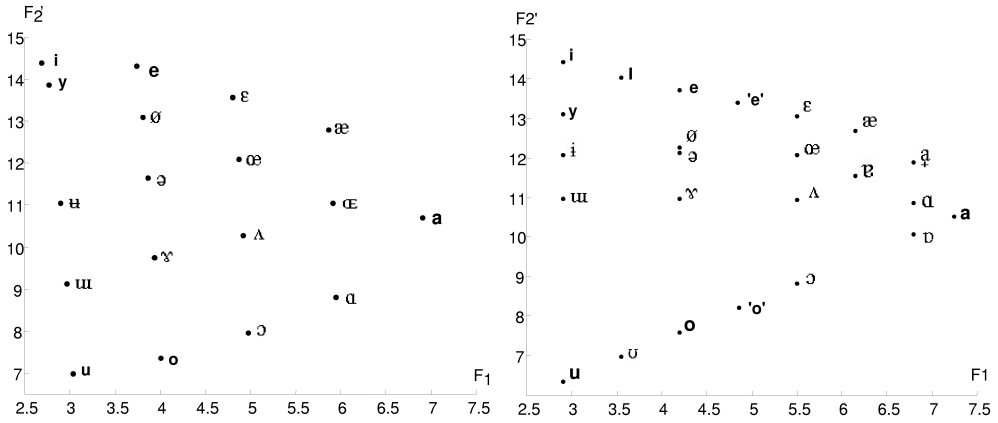

Figure 2
The F_1 and F_2' diagram of prototypical vowels. Left: 18 vowels from INV_L; right: 24 vowels from INV_S.

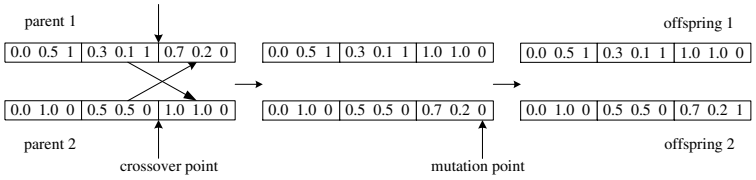

Figure 3
Crossover and mutation operations in simulation of three-vowel systems.

vowels' position as shown in the International Phonetic Alphabet (IPA) vowel chart. Although the IPA vowel chart is better interpreted as an acoustic chart than as an accurate projection of the real articulation of vowels, the chart can still be assumed to reflect the relative positions of different vowels for articulation.

Two inventories of vowel prototypes are used, denoted as INV_L and INV_S, respectively. The former consists of 18 vowels from a set of 19 vowels given in Lindblom (1986).[2] The latter set includes 24 vowels extracted from the set of 33 vowels given in Schwartz et al. (1997a). Figure 2 shows the two inventories in terms of the vowels' first formants (F_1) and transformed second formants (F_2') (Fant 1966) (see Section 3.2 for further explanation), both expressed in terms of the Bark scale (Hartmann 1997).

In the GA model, one-point crossover and one-point mutation are used. Take the simulation of three-vowel systems as an example. Two chromosomes are selected from the population as parents, as shown in Figure 3. Parent 1 includes three vowels: ɤ̈, ø, and ɛ, represented by [0.0 0.5 1], [0.3 0.1 1], and [0.7 0.2 0], respectively. And the three vowels ɯ̟ɪ, ə, and ɑ are included in parent 2, represented by [0.0 1.0 0], [0.5 0.5 0], and [1.0 1.0 0], respectively. Crossover is randomly chosen to take place between two vowels, say, the second and the third vowel in the example, and the two chromosomes exchange their vowels. Next, at random, a mutation occurs in the third vowel in the second offspring: [ɛ] is changed to [œ]. So the two offspring generated from the pair of parents are [ɤ̈, ø, ɑ] and [ɯ̟ɪ, ə, œ].

2 The original inrounding front vowel [y] is deleted, as it is not a common primary vowel, and the outrounding [ü] is changed to symbol [y] to conform with the IPA transcription.

The crossover and mutation rates are both set to 1.0 in order to give GA a high efficiency in searching for the optimal solution. If the genetic operations generate an offspring with the same vowel occurring more than once in a system, this offspring is removed from the population, and a new chromosome is randomly generated. These offspring with higher levels of fitness than those that have been removed are then inserted into the population to keep the population size constant.

3.2 Fitness Evaluation Functions

Two sets of criteria are taken into account in evaluating fitness of the candidate vowel system. One considers only the principle of maximal perceptual contrast (Liljencrants and Lindblom 1972), and the other considers both the intervowel perceptual distance and the intravowel spectral salience related to the proximity of formants (i.e., the dispersion-focalization principle proposed by Schwartz et al. [1997a]).

For the first criterion, the objective is to minimize the following fitness function:

$$\mathcal{F}_1 = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{1}{d_{ij}^2} \tag{1}$$

where d_{ij} is the perceptual distance between vowels i and j. Various metrics for calculating perceptual distance between vowels have been proposed based on perceptual experiments manipulating different combinations of formants and amplitudes of speech signals (Schwartz et al. 1997a). Usually the acoustic parameters, that is, the higher formants (F_2, F_3, and F_4), are first combined and transformed into an "equivalent second formant" F_2', and the auditory distance between two vowels is calculated as the weighted Euclidean distance in the space of F_1 and F_2', where the weight between F_1 and F_2' is determined by λ:

$$d_{ij} = \sqrt{(F_{1i} - F_{1j})^2 + \lambda (F_{2i}' - F_{2j}')^2} \tag{2}$$

In this study, two methods of calculating F_2' are tried, one proposed in Fant (1966), and one given in Schwartz et al. (1997a), to examine the effect of different metrics of calculating perceptual contrast. Therefore, we will have two fitness functions regarding the first criterion, perceptual contrast, denoted as $\mathcal{F}_{1F}$ and $\mathcal{F}_{1S}$, with the latter based on Schwartz's method and the former based on Fant's.

The second criterion includes another objective in addition to the above $\mathcal{F}_1$, the intravowel formant convergence:

$$\mathcal{F}_c = \sum_{i=1}^{n} \frac{-1}{(F_{2i} - F_{1i})^2} + \sum_{i=1}^{n} \frac{-1}{(F_{3i} - F_{2i})^2} + \sum_{i=1}^{n} \frac{-1}{(F_{4i} - F_{3i})^2} \tag{3}$$

The overall fitness function for the second set of criteria is a weighted summation of the above two objectives:

$$\mathcal{F}_2 = \mathcal{F}_1 + \alpha \mathcal{F}_c \tag{4}$$

The values of λ and α are crucial for the prediction. Schwartz et al. (1997a) tested a number of values and found that the following ranges give the best prediction with their vowel inventory:[3] $0.04 \leq \lambda \leq 0.09$ and $0 \leq \alpha \leq 0.4$. In our experiments, we choose the values $\lambda = 0.0625$ and $\alpha = 0.3$, which are within the above ranges.

3 Note that the λ in our formula 2 corresponds to λ^2 in Schwartz et al. (1997a); the range of λ is therefore modified accordingly.

Table 1
Frequent vowel systems in UPSID and predicted vowel systems using different inventories and fitness functions.

N	Observed systems	Predicted systems	
		INV_S	INV_L
3	[i, a, u](14)	$\mathcal{F}_{1F}$: [i, a, u] $\mathcal{F}_{1S}$: [i, a, u] $\mathcal{F}_2$: [i, a, u] S_0: [i, a, u]	[i, a, u] [i, a, u] [i, a, u]
4	[i, 'e', a, u](14) [i, a, u, ɨ](5) [i, a, 'o', u](2) [e, a, o,ɘ](2)	$\mathcal{F}_{1F}$: [i, ɛ, a, u] $\mathcal{F}_{1S}$: [i, ɛ, a, u] $\mathcal{F}_2$: [i, 'e', a, u] S_0: [i, 'e', a, u]	[i, ɛ, a, u] [i, ɛ, a, u] [i, ɛ, a, u]
5	[i, 'e', a, 'o', u](97) [i, ɛ, a, u, ɨ](3)	$\mathcal{F}_{1F}$: [i, ɛ, a, 'o', u] $\mathcal{F}_{1S}$: [i, ɛ, a, 'o', u] $\mathcal{F}_2$: [i, ɛ, a, 'o', u] S_0: [i, 'e', a, 'o', u]	[i, ɛ, a, ɔ, u] [i, ɛ, a, ɔ, u] [i, ɛ, a, ɔ, u]
6	[i, 'e', a, 'o', u, ɘ](26) [i, 'e', a, 'o', u, ɨ](12) [i, 'e', æ, ɪ, 'o', u](12) [i,e, a, ɔ, o, u](4)	$\mathcal{F}_{1F}$: [i, e, æ, a, 'o', u] $\mathcal{F}_{1S}$: [i, e, æ, a, o, u] $\mathcal{F}_2$: [i, e, æ, a, 'o', u] S_0: [i, ɛ, a, 'o', u, ɨ]	[i, ɛ, a, ɔ,u, ɪ̈] [i, æ, a, ɔ, u, ø] [i, ɛ, a, ɔ, u,ɪ̈]
7	[i, e, ɛ, a, ɔ, o, u](23) [i, 'e', æ, a, 'o', u, ɘ](6) [i, 'e', a, 'o', u, ɘ, y](5) [i, 'e', a, 'o', u, ɨ, ɘ](4) [i, e, ɛ, a, 'o', u, ɨ](3)	$\mathcal{F}_{1F}$: [i, 'e', æ, a, 'o', u, ɪ̈] $\mathcal{F}_{1S}$: [i, e, æ, a, ɔ,u, ɪ̈] $\mathcal{F}_2$: [i, e, æ, a, 'o', u, ɪ̈] S_0: [i, e, ɛ, a, 'o', u, ɪ̈]	[i, e, ɛ, æ, a, ɔ, u] [i, e, æ, a, ɔ, u, ɪ̈] [i,e, æ, a, ɔ, u, ɪ̈]

3.3 Results and Analysis

We predict the optimal three- to seven-vowel systems using six sets of experiments, each involving a combination of one of the two vowel inventories (INV_L and INV_S) and one of the three different fitness functions ($\mathcal{F}_{1F}$, $\mathcal{F}_{1S}$ and $\mathcal{F}_2$). The predicted systems are listed in Table 1, together with the commonly observed systems found in the UPSID database, as given in Schwartz et al. (1997a), and the predictions given in Schwartz et al. (1997a) using the same parameters as those for $\mathcal{F}_2$ here (listed as S_0).

First, we compare the predictions resulting from experiments using the same vowel inventory but different fitness functions. The two perceptual distance metrics ($\mathcal{F}_{1F}$ and $\mathcal{F}_{1S}$) produce the same predictions for systems of small sizes (i.e., three-, four- and five-vowel systems), but different predictions for larger systems (i.e., six- and seven-vowel systems), which means that predictions for larger systems are more sensitive to the transformation of F_2'. That transformation used in Schwartz et al. (1997a) produces a more spread perceptual space in the F_2' dimension, and especially [i] has a much larger F_2' than Fant's transformation. It is hard to give an overall evaluation of which perceptual distance metric gives better predictions based on the results we obtained. $\mathcal{F}_{1S}$ predicts a symmetric six-vowel system, whereas $\mathcal{F}_{1F}$ does not, when INV_S is used as the vowel inventory; when INV_L is used as the vowel inventory, however, $\mathcal{F}_{1S}$ predicts a strange six-vowel system with a front rounded vowel ø that is rarely attested in primary vowel systems, whereas $\mathcal{F}_{1F}$ predicts a system close to the second frequent observed system in natural languages.

From the comparison of the two fitness functions $\mathcal{F}_{1S}$ and $\mathcal{F}_2$, we can see that the predictions they make in most cases are the same. There are some exceptions (i.e. in those experiments using INV_S in four-, six- and seven-vowel systems and in

those using INV_L in six-vowel systems). Predictions given by $\mathcal{F}_2$ differ from those given by $\mathcal{F}_{1S}$ only in some small variations between ɛ and 'e', 'o' and o, ɔ and 'o'. $\mathcal{F}_2$ does not predict more nonperipheral vowels than $\mathcal{F}_{1S}$, inconsistent with the proposal of Schwartz et al. (1997a), although the parameters are set within the optimal range suggested by them. The discrepancies between our results and Schwartz et al. (1997a) may need further examination, because our experiment using the same $\mathcal{F}_2$ and INV_S as those in Schwartz et al. (1997a) produced different predictions of optimal five-, six- and seven-vowel systems than those they reported (reproduced as the list of S_0 in the table), given $\lambda = 0.0625$ and $\alpha = 0.3$. When λ is set to 0.09 for the five-vowel system, however, and when $\lambda = 0.025$ and $\alpha = 0.1$ for the six-vowel system, the predictions are the same as those in Schwartz et al. (1997a).

Second, we compare predictions resulting from experiments employing the same fitness function but different inventories. Though the original data set from which INV_S is selected is said to be carefully controlled to sample the acoustic space as evenly as possible (Schwartz et al. 1997a), we do not see much difference between INV_S and INV_L for predicting small-sized systems. If we consider the vowel ɛ in Lindblom's inventory to be equivalent to the vowel 'e' in Schwartz et al. (1997a), ɔ to be equivalent to 'o', and ɨ to be equivalent to ɯ, then the predictions resulting from the two inventories are almost the same. This is not surprising, since we can see from Figure 2 that the peripheral vowels in the two vowel inventories are almost the same in the F_1-F_2' plane. However, for the six-vowel system produced in combination with any of the three fitness functions and the seven-vowel systems produced in combination with the first fitness function ($\mathcal{F}_{1F}$), there are some big differences between the two vowel inventories. This may be mainly because the nonlow unrounded back vowels are much farther away from rounded back vowels in INV_S than in INV_L.

Comparing the predicted systems with the observed systems, we find that only the most frequently observed three- and four-vowel systems are predicted, and that other predictions do not match the observed systems. The reasons for these results may include the following: First, the vowel inventories, especially INV_L, do not provide enough prototypes for predicting vowel systems, such as 'e' and 'o', which are not included in INV_L; second, the perceptual distance metric used in the study may have not perfectly reflected the actual human perception mechanism; and third, only one optimal system can be identified using the current simple GA model. Further considerations of the optimization criteria for the GA model that include the sufficient perceptual contrast principle proposed by Lindblom (1986) instead of the maximal perceptual contrast principle may lead to more consistent predictions. Also the central vowel ə, which occurs often in large vowel systems, may call for another optimization criterion in addition to the current ones (Schwartz et al. 1997a). GA models are promising in carrying out such investigations in which multiple optimization criteria are addressed simultaneously. The following study on the optimal tone systems is such an experiment.

4. Optimization Model for Tone Systems

A tone language is a language having lexically contrastive pitch on each syllable (Pike 1948). Tone languages are found in many parts of the world (Wang 1991). Though tone is as fundamental a constituent in tone languages as are vowels and consonants, to our knowledge, there has been no explanatory or numerical model proposed to enable studies of the universal structure of tone systems comparable to those being conducted for vowel and consonant systems. In this section, we extend the GA models reported above for vowel systems to study the configuration of tone systems from the

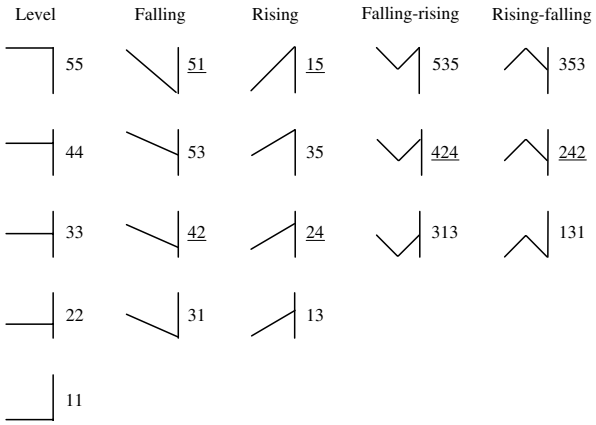

Figure 4
Numeric and corresponding graphic representation of the 19 tones proposed in Gandour (1983). Tones not found in Wang (1967) are underlined.

optimization perspective. Two different sets of criteria are investigated. The first set considers only the perceptual contrast between tones, as in the study of vowel systems in Section 3; the second takes both perceptual contrast and markedness complexity into account. The predicted systems are analyzed and compared with those reported from empirical studies.

4.1 Tone Inventory and Chromosome Representation

In a procedure similar to the one we used for the simulation of vowels, we first choose a tone inventory from which a system selects individual tones. Wang (1967) suggested 13 idealized tones in his study of the phonological features of tones, including five level (11, 22, 33, 44, and 55), two rising (35 and 13), two falling (53 and 31), two falling-rising (535 and 313), and two rising-falling (353 and 131).[4] These 13 tones are considered to represent the maximum contrasts found between tones in any language. In a later tone perception experiment, Gandour (1983) used an extended set of 19 tones, adding to Wang's list two rising tones (15 and 24), two falling tones (51 and 42), and two complex tones (424 and 242). In this study, we take Gandour's 19 tones, shown in Figure 4, as the inventory.

In the model, a chromosome is represented by a number of tones, each of which is selected from among the 19 tones of the inventory. Each tone is described by three numerals representing its shape. The genetic operations and parameters are the same as those in the vowel models.

4.2 Fitness Evaluation Functions

4.2.1 The First Objective. It is assumed that tone systems tend to have a maximum perceptual contrast within the system, much like that proposed for vowel systems. As in the study of vowel systems, we need to derive a method to calculate the perceptual distance between tones. In this study, we use the experimental results from Gandour (1983) to develop such a metric to measure this perceptual distance.

Gandour's experiment was designed to investigate the perceptual dimension of tone and the effect of linguistic experience on a listener's perception of tone. Gandour

4 Tones are represented according to Chao's (1930) conventional five-level transcription system.

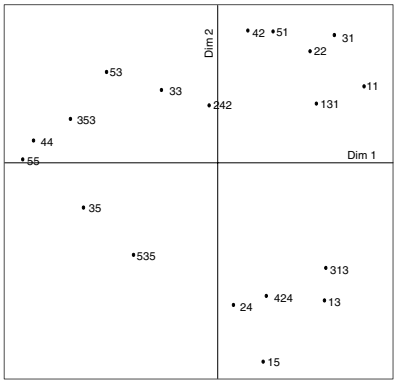

Figure 5
Dimensions 1 and 2 of the two-dimensional INDSCAL tone space (adapted from Gandour [1983]).

synthesized speechlike monosyllables [wa] with the 19 types of tones superimposed on them. Four groups of participants who were from four tone languages (Cantonese, Mandarin, Taiwanese, and Thai) and one group from a nontone language (English) made judgments about the degree of dissimilarity between paired stimulus tones. The collected data were analyzed by an INDSCAL (individual differences scaling) model, and the perceived perceptual dissimilarity among the 19 tone types is presented in a perceptual space in Figure 5. Using these results, we design a metric for calculating the perceptual distance between a pair of tones i and j as their Euclidean distance in the perceptual plane:

$$d_{ij} = \sqrt{(Dim_{1i} - Dim_{1j})^2 + (Dim_{2i} - Dim_{2j})^2} \tag{5}$$

where Dim_{1i} and Dim_{2i} represent the two coordinates of tone i in the perceptual plane. Similarly to the method used in the vowel systems, the perceptual contrast within a tone system is measured by the total perceptual distance for all pairs of tones. The fitness function is therefore the same as the $\mathcal{F}_1$ given in Section 3.2.

4.2.2 The Second Objective. Following the first objective of considering perceptual contrast, the second consideration would naturally be determining the cost of producing various tones. Various mechanisms for controlling the tension of vocal cords and subglottal air pressure and their effect in regulating pitch change have been proposed (Ohala 1978). Different laryngeal muscles, such as the cricothyroid, sternohyoid, and sternothyroid muscles, are found to perform various actions in raising or lowering vocal pitch. An asymmetry is found in the maximum speed of rises as compared to falls in pitch change, which suggests that falling tones may in some sense be easier to produce than rising tones (Collier 1984; Ohala 1978). Level tones are found to be universally preferred to contour tones, and simple contour tones to complex contour tones (Maddieson 1978), which may be due to the different production cost of various types of tones. It is hard, however, to quantify these differences, and no systematic measurements have been made available yet.

Because of the lack of data measuring the effort required to produce different tones, we chose another criterion as the second objective: the markedness complexity, which is based on a study of phonological features of tones proposed in Wang (1967). Each tone is assigned a complexity value based on the analysis of tones with features

Table 2
Relative complexity of tones as defined by marking conventions (adapted from Wang [1967]).

	55	11	44	22	33	35/15	13/24	53/51	31/42	535	313/424	353	131/242
CONTOUR	u	u	u	u	u	m	m	m	m	m	m	m	m
HIGH	+	−	+	−	−	+	−	+	−	+	−	+	−
CENTRAL	u	u	m	m	m	u	u	u	u	u	u	u	u
MID	u	u	u	u	m	u	u	u	u	u	u	u	u
RISING	u	u	u	u	u	+	+	−	−	+	+	+	+
FALLING	u	u	u	u	u	−	−	+	+	+	+	+	+
CONVEX	u	u	u	u	u	u	u	u	u	u	u	m	m
COMPLEXITY	1	1	2	2	3	4	4	4	4	4	4	5	5

and marking conventions, as shown in Table 2. In the table, "m" stands for the marked specification, which is the favored specification, and "u" for unmarked. As there is no empirical ground for favoring either +HIGH or −HIGH, or RISING or FALLING, the features of [HIGH], [RISING] and [FALLING] are only analyzed by its presence or not by assigning + or −, instead of specifying their markedness. We note that the assignment of + and − may need further justification or modification with regard to our discussion above on rising and falling tones. In this study, however, we still adopt this analysis presented in the table for our simulation. The specifications "m," "+," and "−" each add one unit to the complexity of a particular tone, whereas "u" does not.

Although we have maintained in Table 2 the original complexity assignment for the 13 typical tones given in Wang (1967), we have incorporated the six additional tones proposed by Gandour (1983). It is assumed that tone pairs such as 35 and 15, 13 and 24, 53 and 51, 31 and 42, 313 and 424, and 131 and 242 are of the same complexity. As far as we are aware, there is no tone system having more than two falling or rising contrasts that all provide lexical distinctions. Those transcriptions of tone systems that do incorporate more than two rising or two falling tones may be due to overdifferentiation within a single tone paradigm, as suggested in Wang (1967).

Markedness is a method of representing the linguist's knowledge of a phonological system. This knowledge derives primarily from observations of three sorts: the frequency of distribution of the sounds in the languages of the world, the patterns of historical change in sound systems, and the acquisition of sounds in children and the dissolution of sounds in linguistic pathology. Therefore the complexity assigned to tones based on markedness may reflect an integrated effect of perception, production, and learnability.

The two objectives, perceptual distance and markedness complexity, are taken into account in a MOGA model using the Pareto ranking method, as introduced in Section 2.2, to predict optimal tone systems. Simulation results are shown in Section 4.4.

4.3 Empirical Data Analysis

Before reporting the results of our simulation, we report our analysis of an available database of observed tone systems with which we can compare our predictions. To our knowledge, there is no large database for tone systems of the same scale as SPA or UPSID for vowel and consonant systems. We have been able to find, however, a computer database consisting of 737 entries from various dialect[5] locations in China

5 The "dialects" we used here are referred to as the various languages such as Mandarin, Cantonese, Min dialect, and Wu dialect.

Table 3
Frequencies of occurrence of the 19 tone types in the normalized Cheng database.

Tone	31	53	55	42	35	24	313	44	33	13
Frequency	326	294	291	281	262	248	241	219	185	167
Tone	11	51	22	424	353	242	131	15	535	
Frequency	86	78	66	27	12	8	8	6	1	

Table 4
Frequencies of occurrence and different sizes of tone system in both the original and the normalized Cheng database.

Size	3	4	5	6	7	8	9	Total
Frequency(original)	26	497	105	91	10	7	1	737
Frequency(normalized)	22	448	86	79	4	2		641
Types(normalized)	19	162	72	60	4	2		319

compiled by Cheng (1973). Some of the 737 entries are from the same location, but from different reports or at different times. They are considered individual systems in our analysis since they were obtained independently and we want to consider as many systems as possible in our analysis, though it is possible that our analysis may be contaminated by the unequal quality of the entries.

To make comparison with predicted systems possible, we normalize nontypical tones in Cheng's database to the 19 typical ones. For example, tones 54, 12, and 324 are converted to 53, 13, and 313, respectively. Such normalization also provides some advantage in dealing with different types of transcription of the same system. Very often different researchers have their own individual strategies of transcription, and different informants employed for the same dialect also have individual differences in their speech. Therefore controversies often arise, such as whether a particular tone in Cantonese is 13 or 23. By normalizing the tones in Cheng's database into the 19 typical tones, some such controversies will be resolved. Table 3 shows the frequencies of occurrence of the 19 types of tones in the normalized database. When the changes in the database as a result of normalization create a situation in which the same tone occurs more than once in one system, that system is excluded from the normalized database for further analysis. The original database includes 737 systems with 606 different tone types. After normalization, there were 641 systems of 319 different tone types. Table 4 shows the frequencies of different sizes of tone system in both the original and the normalized database. Systems with four tones are by far the most frequent type.

Figure 6 shows the tone types occurring in the four-tone normalized systems. Frequencies of occurrence are indicated beside each tone. We can see that tones are heavily clustered in three areas in the perceptual plane: the upper right (including tones 42, 51, 31, 22, 11, 131) (cluster 1), the lower right (24, 13, 424, 313, 15) (cluster 2), and the middle left (55, 44, 35, 53, 33, 353, 242) (cluster 3). Tone 535, although it does not occur in the normalized four-tone systems, can also be classified in cluster 3, as inferred from Figure 5. When we calculate the frequencies of occurrence of tones in these three areas for three-, four-, and five-tone systems, as shown in Table 5, we find that most of the systems tend to select tones from each of the three individual clusters,

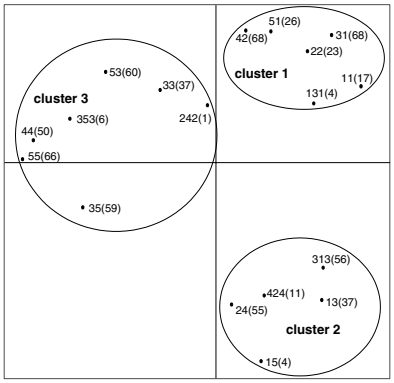

Figure 6
Frequency of tone occurrence in observed four-tone systems shown in the perceptual space (total 162 systems).

Table 5
Number of tones in the three clusters in perceptual space and percentage of systems having tones from each of the three clusters.

	Number of systems	Number of tones in C1	Number of tones in C2	Number of tones in C3	Percentage of systems having tones from each of the three clusters
Three-tone system	19	19	17	21	73.7
Four-tone system	162	206	163	279	75.0
Five-tone system	72	107	91	162	74.3

rather than selecting tones from only one or two clusters. The percentages of systems having tones from each of the three clusters are shown in the last column of Table 5. We can infer from the data in Table 5 that there tends to be a great deal of perceptual contrast within the systems, since the tones from separate clusters often have larger perceptual distances between them than tones in the same cluster. Furthermore, four- and five-tone systems do not employ tones from the three clusters evenly. It is obvious that cluster 2, which includes most of the rising tones (except high rising tone 35, which is contoured in cluster 3), contributes fewer tones to four- and five-tone systems than the other two clusters. This implies that rising tones are less preferred in the observed systems.

4.4 Simulation Results and Comparison with Observed Systems

Tables 6, 7, and 8 give the predicted "optimal" three-, four- and five-tone systems obtained from the GA model. A number of tone systems that occur frequently in the normalized database are also given in the tables for comparison. The frequencies with which the tone systems were observed in the normalized database are given in parentheses besides the tones.

For a system of a given size, the model using only the first objective (maximal perceptual contrast) predicts only one optimal system, which is indicated with asterisks in the tables. This optimal system has the maximum perceptual contrast but the largest markedness complexity. On the other hand, the two-objective model predicts a number of optimal systems besides the optimal one predicted by the single-objective

Table 6
Predicted optimal and frequent observed three-tone systems.

Predicted systems					Observed systems				
Perceptual distance	Complexity	t1	t2	t3	Perceptual distance	Complexity	t1	t2	t3 (frequency)
*14.41	9	55	31	15	12.97	9	55	31	424(3)
13.77	7	55	22	15	11.52	12	42	35	13(1)
13.97	6	55	11	31	8.84	11	33	53	313(2)

Table 7
Predicted optimal and frequent observed four-tone systems.

Predicted systems						Observed systems					
Perceptual distance	Complexity	t1	t2	t3	t4	Perceptual distance	Complexity	t1	t2	t3	t4 (frequency)
*24.63	12	55	11	31	15	22.42	14	44	53	31	13(21)
24.55	10	55	11	31	15	22.02	13	55	42	31	24(27)
23.67	9	44	22	11	15	21.98	13	55	53	31	24(17)
23.24	8	55	44	11	13	21.03	13	55	51	35	313(20)

Table 8
Predicted optimal and frequent observed five-tone systems.

Predicted systems							Observed systems						
Perceptual distance	Complexity	t1	t2	t3	t4	t5	Perceptual distance	Complexity	t1	t2	t3	t4	t5 (frequency)
*39.20	17	55	53	31	15	13	35.90	13	55	44	22	31	13(1)
38.44	14	55	11	53	15	13	35.82	17	55	42	31	24	313(4)
38.60	12	55	44	11	31	15	34.54	15	55	44	42	24	13(1)
38.17	10	55	44	22	11	15	32.33	12	55	22	11	31	24(1)
							32.15	10	55	44	33	11	24(1)
							28.12	17	11	51	53	42	24(2)
							27.09	18	44	53	42	31	35(3)

model. These optimal systems, like the points A, B, C, and D shown in Figure 1, form a Pareto-optimal set. They are equally good in terms of the overall fitness including the two objectives (large perceptual contrast and small markedness complexity). The two-objective model may be viewed as a possible implementation of the sufficient perceptual contrast proposed for the study of optimal vowel systems by Lindblom (1986).

We can observe in the predicted systems characteristics similar to those in the observed systems, particularly the uneven utilization of the three clusters in the perceptual space. If we assume that tones within the same cluster are interchangeable, then we can see that most of the predicted optimal systems have corresponding observed systems, except one prediction for a three-tone system (55, 11, 31), which does not include a tone from cluster 2.

The predicted systems seem to exploit more tones in the outer area of the perceptual space, such as 55, 11, 31, and 15, whereas in the observed systems, tones in

the relative inner area, such as 35, 53, 42, 24, and 313, are more frequent. High rising tone 15 occurs often in predicted systems because of its high salience in the perceptual space. In the observed systems, however, tone 15 is very infrequent (only six occurrences among the 641 systems observed). In the observed systems, pairs of contrasting tones, such as 24 and 42, 13 and 31, occur quite often. These contrasting pairs are not clear in our prediction. Tone 31 co-occurs more frequently with tone 15 than with tone 13, which again is due to the perceptual salience of tone 15.

The observed systems include more tones close to the center of the perceptual space, whereas the predicted systems prefer tones located in its periphery. This pattern is similar to one observed in a long-standing problem in the study of vowel systems (Liljencrants and Lindblom 1972). In observed vowel systems, especially the larger systems, central vowels commonly occur (examples can be seen in Table 1), whereas the proposed optimal vowel systems predict them only rarely. The utilization of fewer peripheral areas in observed systems than in predicted ones suggests that the role of maximizing perception contrast in the models may need to be adjusted or that more optimization criteria in addition to the perceptual contrast and markedness complexity should be added.

Moreover, the markedness complexity we hypothesize is an abstract measure that incorporates many factors, including perception and production. Thus the consideration of perception contrast may have been duplicated in the fitness function. This may be another reason for the discrepancy mentioned in the preceding paragraph. When more empirical studies in the physiology of tone production are available, we may consider tone production as an individual objective in the fitness function, which we would expect to allow better predictions.

5. Conclusions and Discussion

In this article, we apply optimization models using GAs to study the configuration of vowels and tone systems. The approach we use is similar to that in previous explanatory models that have been used to study vowel systems. Certain criteria, which are assumed to be the principles governing the structure of sound systems, are used to predict optimal systems. In most of the previous studies (Liljencrants and Lindblom 1972; Crothers 1978; Lindblom 1986), only one criterion has been considered. When two criteria have been considered, the two objectives are combined into a single weighted function (Boë, Schwartz, and Vallée 1994). In our study of vowel systems, the simple GA model we use also adopts a weighted function to combine two criteria, perceptual contrast and focalization. In our study of tone systems, however, we apply a MOGA model that uses a Pareto ranking method to consider two criteria, perceptual contrast and markedness complexity, simultaneously, without combining them into a scalar function. A priori knowledge of the weights of the two criteria are not necessary.

Another advantage of an MOGA is that we can obtain a set of Pareto-optimal results, instead of only one. An MOGA model generates more optimal predictions than a single-objective model, and therefore it is more likely to predict more systems that are close to the systems actually observed. Although the consistency between the predicted systems and the observed systems in the current study is not as significant as that obtained for vowel systems, further investigation along this line is promising.

Following the deductive approach pursued in this study, we can design various criteria to predict optimal systems. The deductive approach provides convenience and freedom in the manipulation of different parameters in the models, such as the param-

eters λ and α in Schwartz et al. (1997a), to test different hypothesized mechanisms. It is necessary, however, to seek explanations for such parameters in terms of physiological or cognitive constraints.

Studies taking the deductive approach must not be pursued independent of the inductive approach. For example, in the study of tone systems, few comprehensive tone databases are available. The resources on which our investigation of tone systems relies, including the experiment in Gandour (1983) and the database in Cheng (1973), are based mostly on the observation of tone languages found in Asian. The incorporation of data from other types of tone languages in Africa and America is expected to help in refining our explanatory hypothesis about the configuration of the systems.

Lastly, we would like to point out that although in this study we apply optimization to predict vowel and tone systems, we do not imply that there exist any explicit and/or global optimization processes in the formation of such systems. We have no grounds to believe that speakers are aware of what sounds will provide maximal perceptual contrast or require the least production effort and therefore deliberately choose those sounds. Optimization must be an emergent property from the interactions of language users (de Boer 2000, 2001). Each individual speaker has certain physiological and cognitive constraints which limit the sounds it is possible for him to make and assign preference to certain of those sounds over others. These constraints, however, provide only a range of possible sounds. It is the interactions among individuals that determine precisely which systems among those that are possible will emerge. This is why different configurations of sound systems, even suboptimal ones in the sense of some hypothesized criteria, can be observed in real systems. Research including modeling from this perspective is promising and may lead to more realistic predictions of sound systems.

Acknowledgments

This research is supported in part by two grants from the City University of Hong Kong, nos. 7100096 and 9010001. The second author is also supported by a grant from the Ministry of Education, Science, Sports and Culture of Japan, no. 11610512. We thank C. C. Cheng for providing us with his tone database of Chinese dialects. We are thankful to Lisa Husmann and James Minett for their kind help in preparing this article. Also we greatly appreciate the three reviewers for their very helpful comments and suggestions.

References

Boë, Louis-Jean, Jean-Luc Schwartz, and Nathalie Vallée. 1994. The prediction of vowel systems: Perceptual contrast and stability. In Eric Keller, editor, *Fundamentals of Speech Synthesis and Speech Recognition*. John Wiley, Chichester, England, pages 185–213.

Chao, Yuan Ren. 1930. A system of tone letters. *Le Maître Phonétique*, 45:24–27.

Cheng, Chin Chuan. 1973. A quantitative study of Chinese tones. *Journal of Chinese Linguistics*, 1(1):93–110.

Collier, Rene. 1984. Some physiological and perceptual constraints on tonal systems. In Bernard Comrie, Brian Butterworth, and Östen Dahl, editors, *Explanations for Language Universals*. Walter de Gruyter, Berlin, pages 237–247.

Crothers, John. 1978. Typology and universals of vowel systems. In J. H. Greenberg, editor, *Universals of Human Language (Phonology)*, volume 2. Stanford University Press, Stanford, California, pages 93–152.

de Boer, Bart. 1997. Generating vowel systems in a population of agents. In Phil Husbands and Inman Harvey, editors, *Fourth European Conference on Artificial Life*. Cambridge, MIT Press, pages 503–510.

de Boer, Bart. 2000. Self-organization in vowel systems. *Journal of Phonetics*, 28(4):441–465.

de Boer, Bart. 2001. *The Origins of Vowel Systems*. Oxford University Press, Oxford and New York.

Fant, Gunnar. 1966. A note on vocal tract size factors and non-uniform f-pattern scalings. *Speech Transmission Laboratory Quarterly Progress and Status Report*,

1:22–30.
Fonseca, Carlos M. and Peter J. Fleming. 1998. Multiobjective optimization and multiple constraint handling with evolutionary algorithms—Part I: A unified formulation. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, 28(1):26–37.
Gandour, Jack. 1983. Tone perception in Far Eastern languages. *Journal of Phonetics*, 11:149–175.
Goldberg, David E. 1989. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, Reading, Massachusetts.
Hartmann, William M. 1997. *Signals, Sound, and Sensation*. AIP Press, New York.
Holland, John H. 1975. *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor.
Jakobson, Roman. 1941. *Kindersprache, Aphasie und allgemeine Lautgesetze*. Uppsala. Reprinted in *Selected Writings I*. Mouton, The Hague, 1962, pages 328–401.
Ladefoged, Peter and Ian Maddieson. 1990. Vowels of the world's languages. *Journal of Phonetics*, 18:93–122.
Ladefoged, Peter and Ian Maddieson. 1996. *The Sounds of the World's Languages*. Blackwell, Oxford.
Liljencrants, Johan and Björn Lindblom. 1972. Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, 48:839–862.
Lindblom, Björn. 1986. Phonetic universals in vowel systems. In John J. Ohala and Jeri J. Jaeger, editors, *Experimental Phonology*. Academic Press, Orlando, Florida, pages 13–44.
Lindblom, Björn and Ian Maddieson. 1988. Phonetic universals in consonant systems. In Larry M. Hyman and Charles N. Li, editors, *Language, Speech and Mind: Studies in Honour of Victoria A. Fromkin*. Routledge, London, pages 62–78.
Maddieson, Ian. 1978. Universals of tone. In Joseph H. Greenberg, editor, *Universals of Human Language (Phonology)*, volume 2. Stanford University Press, Stanford, California, pages 335–365.
Maddieson, Ian. 1984. *Patterns of Sounds*. Cambridge University Press, Cambridge.
Maddieson, Ian and Kristin Precoda. 1990. Updating upsid. *UCLA Working Papers in Phonetics*, 74:104–111.
Ohala, John J. 1978. Production of tone. In Victoria A. Fromkin, editor, *Tone: A Linguistic Survey*. Academic Press, New York, pages 5–39.
Pike, Kenneth. 1948. *Tone Languages*. University of Michigan Press, Ann Arbor.
Redford, Melissa A., Chun Chi Chen, and Risto Miikkulainen. 2001. Constrained emergence of universals and variation in syllable systems. *Language and Speech*, 44:27–56.
Schwartz, Jean-Luc, Louis-Jean Boë, Nathalie Vallée, and Christian Abry. 1997a. The dispersion-focalization theory of vowel systems. *Journal of Phonetics*, 25:255–286.
Schwartz, Jean-Luc, Louis-Jean Boë, Nathalie Vallée, and Christian Abry. 1997b. Major trends in vowel system inventories. *Journal of Phonetics*, 25:233–253.
Stadler, Wolfram. 1988. *Multicriteria Optimization in Engineering and in the Sciences*. Plenum, New York.
Van Veldhuizen, David A. and Gary B. Lamont. 2000. Multiobjective evolutionary algorithms: Analyzing the state-of-the-art. *Evolutionary Computation*, 8(2):125–147.
Vihman, Marilyn. 1977. *A Reference Manual and User's Guide for the Stanford Phonology Archive*. Part I. Stanford University.
Wang, William S.-Y. 1967. Phonological features of tone. *International Journal of American Linguistics*, 33:93–105.
Wang, William S.-Y. 1968. The basis of speech. Project on Linguistic Analysis Reports, University of California at Berkeley. Reprinted in *The Learning of Language*, ed. by C. E. Reed, 1971.
Wang, William S.-Y. 1976. Language change. *Annals of the New York Academy of Science*, 280:61–72.
Wang, William S.-Y. 1991. Tone languages. In Kirsten Malmkjaer, editor, *The Linguistic Encyclopedia*. Routledge, London, pages 1455–1470.