Journal of Universal Language 4
September 2003, 117-145

Universal Perceptual Attributes for Perception of American English Vowels by English and Japanese Native Speakers and Implications for Language Typology

Ashuboda Marasinghe,[1] Stephen Lambacher,[1] William Martens,[2] Michael Cohen,[1] Charith Giragama,[1] Susantha Herath[3] & Garry Molholt[4]

University of Aizu[1], McGill University[2], St. Cloud State University[3], West Chester University[4]

Abstract

A universal perceptual space for 10 American English vowel sounds was derived for two groups of listeners–a group of native speakers of English and a group of native speakers of Japanese. Subsets of these two groups made ratings on 12 bipolar adjective scales for the same set of sounds, each of the two groups using anchoring adjectives taken from their native language. Although there was no evidence of any difference between the two groups in their *INDSCAL*-derived perceptual dimensions for these vowel sounds, the adjectives were used differently in describing those same perceptual dimensions by the two groups. Though a few of the adjectives were used to describe similar perceptual variations, language typological implications of this investigation is that caution be exercised in generalizing semantic differential ratings obtained in one language, especially when those ratings are intended to aid in the interpretation of data from listeners speaking a different native language.

Keywords: universal perceptual space, language typology, and semantic differential analysis

1. Introduction

Recent comparative studies in psycholinguistics make it clear that although languages display superficial uniquenesses in phonetics and semantics which render them mutually unintelligible, at a deeper level they display certain universals which render them mutually translatable (Osgood 1975). Universals are found in the limited stock of phonetic features from which each human language draws its phonemes, in the deep cognitive structure of perceptions from which all humans construct sentential propositions about their world, and even in the sets of semantic features with each concept meanings are differentiated.

It is largely unknown how language learners manage to differentiate the vowels of the language to which they are exposed (Kewley-Port 1989). There are at least as many meanings of "meaning" as there are disciplines which deal with the perception of vowel sounds and, of course, many more than this because exponents within disciplines do not always agree with one another. Nevertheless, definitions do tend to correspond generally with the purposes and techniques of the individual doing the defining, focusing on that aspect of the phenomena which his discipline equips her to consider. Thus a sociologist or anthropologist typically defines the meaning of a sound in terms of the common features of the situations in which it is used and of the activities which it produces (Osgood 1957). Of all the processes that constitute the nervous system—that "little black box" in psychological theorizing-the one we call "perception" is held by common consent to be the most elusive. Yet, again by common agreement among social scientists, this system is one of the most important determinants of human behavior. It therefore behooves us to try, at least, to find some kind of universal objective index. To characterize anything that goes on within "the little black box" it is necessary to use some observable output from it as an index. To state the problem in yet another way, we wish to find a kind of measurable activity or behavior of sign-using organisms which is maximally dependent upon and sensitive to meaningful states, and minimally dependent upon perception. These problems include improving speech coding, determining metrics for speech recognition, and understanding the process of dialect variation and language change.

In contrast, a semantic approach to describing the perception of complex sounds has been pursued for more than 40 years (Solomon 1958). This approach is based upon the idea that the way in which words are used can be quantified via **S**emantic **D**ifferential **A**nalysis (***SDA***), a well-established method for the measurement of word meanings using bipolar scales with adjectives of opposite meaning anchoring ends of each scale (Osgood 1957). A fundamental problem in using *SDA* alone to study timbre is that the magnitude of perceptual variation associated with ratings on one semantic scale, relative to other semantic scales, is difficult to determine from such bipolar adjective ratings. An additional practical problem in using *SDA* to study perception is how to deal with language differences— that is, how to relate results obtained in otherwise similar studies that used semantic scales formed from adjectives taken from different languages.

The primary motivation for the present study was strictly methodological, focusing upon the question of whether multilingual semantic scales may be related to a universal perceptual space. Another motivation for this study was the refinement of adjective scales for use in subsequent studies of perception using groups of listeners with various native languages. Two listening experiments were executed to establish both perceptual and semantic scale values for each of 10 American English vowel sounds. In the first experiment, English and Japanese listeners gave dissimilarity ratings for all pairwise comparisons of the vowel sound stimuli. Submitting these obtained dissimilarity ratings to **IN**dividual **D**ifferences **SCAL**ing (***INDSCAL***) analysis yields a universal perceptual space for the stimuli. In the second experiment, the English and Japanese listeners made ratings on a set of 12 bipolar adjective scales for each of the 10 stimuli. It was not assumed a priori outset that ratings on these 12 adjective scales should necessarily capture the most salient differences between the stimuli. The experimental data from the second experiment was analyzed through Principle Component Analysis. These results are also considered in the framework of language typology. Finally, it was hoped that the results would provide a better understanding of how adjectives from each language are used by listeners to describe the sound of American English Vowels as part of separate ongoing research project.

1.1. INdividual Differences SCALing (*INDSCAL*)

The purpose of *INDSCAL* is to represent objects whose dissimilarities are given as points in a metrical space. The distances in the space should be in accordance with the dissimilarities as well as is possible. Besides the configuration, a salience matrix is calculated. Individual subjects may differ in how they form judgments of global dissimilarity, and so a refined method for doing a weighted *INDSCAL* analysis (Borg 1997, Shepard 1972) that takes such individual differences into account is to be recommended. *INDSCAL* analysis is a powerful means for deriving an interpretable representation of the dimensions underlying reported inter-stimulus dissimilarities obtained from a potentially inhomogeneous group of subjects, each of which may place different weights upon each of the perceptual dimensions.

1.2. Principal Components Analysis (*PCA*)

PCA is a well–known statistical method used for essence of information from a vast amount of data obtained from subjective evaluation experiments (Kister 1991, Martens 1987). Subjective evaluation results contain somewhat redundant data, and *PCA* is useful in order to reduce the driving forces governing the redundancy. *PCA* is a method of transforming a number of variables into one or a few linearly independent representative variables (principal components) with the least amount of information lost. Each principal component is a linear combination of the original variables, and all principle components are orthogonal to each other. *PCA* is often confused with Factor Analysis (*FA*) because they share the same goal of replacing a large set of observed variables with a smaller set of new variables (Martens 2002a). However, *PCA* is used for reducing a number of variables to explain the overview of observed data, while *FA* is used for discovering a set of latent (unobservable) variables underlying subjective judgments in adjective ratings.

2. Methods

2.1. Stimuli

Ten American English (AE) vowel sounds were digitally synthesized (Table 1). Similar sounds have been used in speech science ever since Peterson and Barney first published such vowel formant data in 1952.

Table 1

IPA symbol	ASCII code	Example word	Vocal Articulation		
			lips	tongue	mouth
i	EE	bead	unrounded	front	closed
ɪ	IH	bid	unrounded	front	near-closed
ɛ	EH	bed	unrounded	mid-front	open
æ	AE	bad	unrounded	front	near-open
ɑ	AH	bod(y)	unrounded	back	open
ʌ	UH	bud	unrounded	mid-back	open
ɔ	AW	bawd	rounded	mid-back	open
ʊ	OO	bud(dhist)	rounded	near-back	near-closed
u	UU	booed	rounded	back	closed
ɚ	ER	bird	rounded	back	closed

10 Stimuli—Vocal articulation for production of 10 American English vowels (those investigated in Peterson and Barney 1952), including for each vowel its IPA symbol, a code of ASCII characters used for graphics, and an example word (Ladefoged 2001).

2.2. Subjects

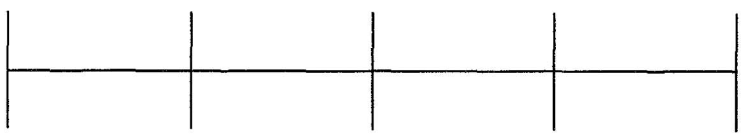
47 Native speakers of English were recruited for the dissimilarity rating task and 28 Native speakers of English were recruited for the semantic differential rating task. All the subjects were undergraduates at West Chester University in Pennsylvania, USA. In addition, 46 native speakers of Japanese were recruited for the Dissimilarity Rating Task and 52 native speakers of Japanese were recruited for the Semantic Differential Rating Task. All the subjects were undergraduates at the University of Aizu in Fukushima Prefecture, Japan.

2.3. Listening Tasks

The two listening experiments were administered to each of the listeners. Directions for the two tasks were presented to the listeners just before each task was performed. All of the instructions were given to each group of listeners in their native language.

2.3.1. Dissimilarity Rating Task

Listeners were instructed to indicate global dissimilarity ratings on a 5-point scale for 90 pairwise comparisons of the American English vowel sound stimuli. Figure 1 shows the 5-point dissimilarity ratings scale. The listeners were asked to listen to each stimulus pair once and rate their global dissimilarity without respect to any particular property. A response of '1' implied that the two samples were perceived as "almost exactly the same" and a response of '5' implied that the two stimuli were perceived as "almost completely different." Each pair of stimuli was presented twice, in a random order, with a 1-second inter-stimulus interval and a 5-second inter-trial interval (the time between each presented pair of stimuli during which responses were recorded).



1	2	3	4	5
exactly same	slightly same	neither same nor different	slightly different	completely different

Figure 1. The 5-point dissimilarity rating scale.

2.3.2. Semantic Differential Rating Task

Listeners were asked to rate each stimulus on 12 bipolar adjective scales. Again, a 5-point scale was employed, but in this case, a response of '1' indicated that the vowel stimulus was best characterized by the adjective anchoring one end of the semantic differential scale, while a response of '5' indicated that the stimulus was best characterized by the adjective anchoring the other end of the scale. Figure 2 shows such a 5-point bipolar adjective scale. Listeners were instructed to give a response of '3' if neither of the anchoring adjectives characterized the stimulus. The 10 stimuli were presented in a order for each of 12 adjective pairs, with a 5-second interval between each individual stimulus presentation. Adjective scales were given to listeners in their native language. Table 2 shows the corresponding bipolar adjective pairs in two languages, listed in order of presentation for subsequent semantic differential ratings.

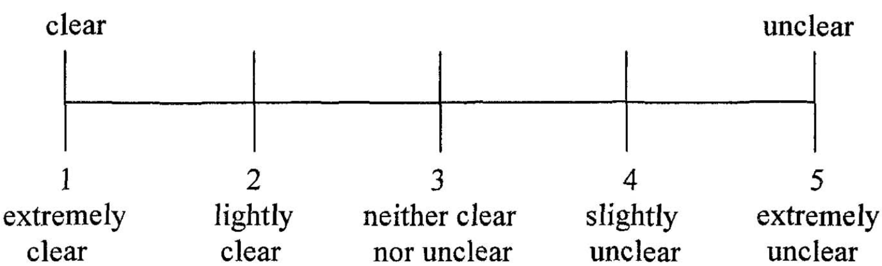


Figure 2. The 5-point bipolar adjective scale. The clear-unclear bipolar adjective pair is used as an example. The same kind of scale was used for all 12 adjective pairs.

The working assumption in this study is that complex sounds having multiple perceptual attributes have a mental structure that can be quantitatively captured in terms of a multidimensional perceptual space that is distinct from the words that might be used to describe the individual perceptions occupying that space. It is hypothesized that the dimensions of perceptual space for a small set of stimuli may be common among groups of listeners with differing native languages. It is further hypothesized that the words used by multilingual groups of listeners may share common underlying semantic structures when used to describe that small set of stimuli. Determining whether or not either of these hypotheses can be supported by experimental data is the primary goal of this study.

Table 2

Code	Japanese	English
CLER	hakkiri-shita – kumotta	clear – unclear
SHRP	surudoi – nibui	sharp – dull
BRIT	akarui –kurai	bright – dark
HIGH	takai – hikui	high – low
HEAV	omoi – karui	heavy – light
TRAN	sunda – nigotta	transparent – muddy
COMP	konpakuto na – hirogatta	compact – diffuse
CALM	ochitsuita – sozoshii	calm – clamorous
SMOO	nameraka na – arai	smooth – rough
THCK	atsui – usui	thick – thin
MAGN	dodotoshita – hikaemena	magnificent – humble
POWR	hakuryokunoaru – yowai	powerful – weak

Corresponding bipolar adjective pairs in two languages. The Japanese adjectives were translated from English by a native speaker of Japanese (Martens 2000). A four-lettered Code is used to key figures 7 and 8.

This research, then, employed both **M**ulti **D**imensional **S**caling (***MDS***)– and *SDA*–based methods in a study of the perceptual variation associated with vowel sounds. Both of these methods have limitations which are largely circumvented when they are combined; e.g., via joint analysis (Ramsy 1980). One limitation of *SDA*–based methods, however, is not circumvented via combination with *MDS*–based methods, and this limitation had to be addressed in a preliminary experiment regarding adjective selection (Shiffman 1981). The power of *SDA* rests upon good choice of the bipolar adjective scales on which listeners are to rate each of the stimuli, and so special attention was given to this stage of the study. This relatively bias free means for adjective selection is regarded as an essential methodological detail of any study that explores a relatively unknown stimulus domain using *SDA*.

3. Results and Discussion

3.1. Dissimilarity Ratings Results

Ratings of dissimilarity were reported for all pairwise comparisons of the 10 American English vowel stimuli native speakers of English and Japanese. Each stimulus pair was presented only once, and listeners used a 5-point scale for their ratings.

A separate 10×10 matrix of dissimilarity data was constructed for each of the listeners, and these were combined into a single submission for *INDSCAL* analysis, using the *ALSCAL* routine found in the *SPSS* statistical analysis software.

Figure 3 shows the *INDSCAL*–derived subject space based upon the dissimilarity judgments. Plotted here are the weights placed by each subject on each dimension in generating their dissimilarity ratings. Though individual subjects placed slightly different weights on each of the two dimensions in this result, each stimulus is located at a single pair of coordinate values in the derived subject space that fit best the responses of all listeners.

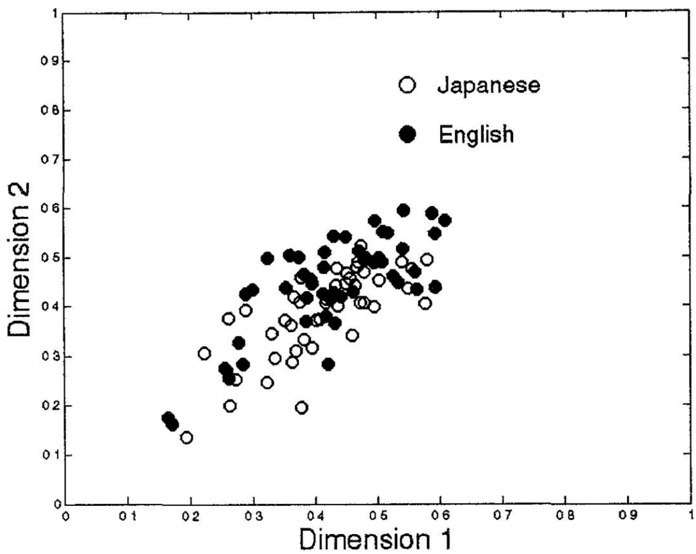


Figure 3. *INDSCAL*–derived subject space for 47 native English and 46 native Japanese speakers

3.1.1. Universal Perceptual Space for Vowel Sounds

One way to determine whether two groups of listeners share a single, universal perceptual space is to examine the ways in which they generate dissimilarity judgments for pairwise comparison of stimuli taken from the set of perceptions of interest. Individual subjects may differ in how they form judgments of global dissimilarity, and so a refined method for doing a *weighted MDS* analysis (Martens 2002a, Tucker 1963) that takes such individual differences into account is to be recommended. This article discusses the use of *INDSCAL* analysis (Martens 2002b) as a powerful means for deriving an interpretable representation of the dimensions underlying reported inter-stimulus dissimilarities obtained from a potentially inhomogeneous group of subjects, each of whom may place different *weights* upon each of the perceptual dimensions. While sets of dissimilarity data can be averaged across subjects in each group to obtain two aggregated datasets for submission to classical *MDS* analysis through *INDSCAL*, such dissimilarities shows the advantages provided by the *INDSCAL* model for the analysis of multiple sets of dissimilarity data, without requiring the assumption of a homogeneous group of subjects who share an identical perceptual structure for the stimuli. Figures 4 and5 show the *INDSCAL*–derived perceptual spaces for the two groups of listeners.

According to the correlation shown in Table 3, English and Japanese native speakers share a universal perceptual space, as shown in Figure 6. The overlap between the dimensional weights for native speakers of English and Japanese also indicates (Figure 3) that the two groups are in general agreement on the stimulus dimensions and their relative importance.

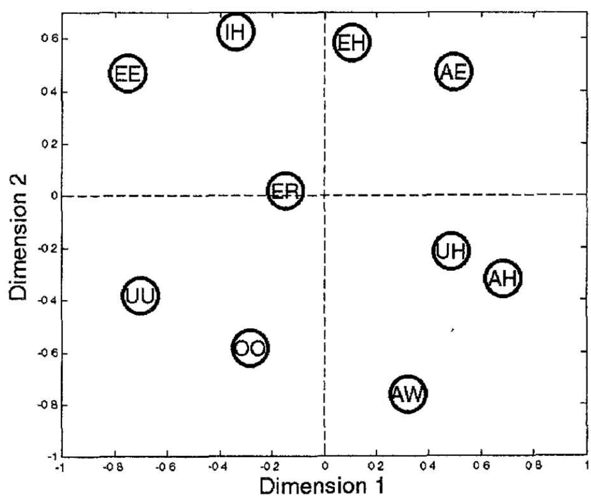


Figure 4. *INDSCAL*–derived perceptual space for native English speakers.

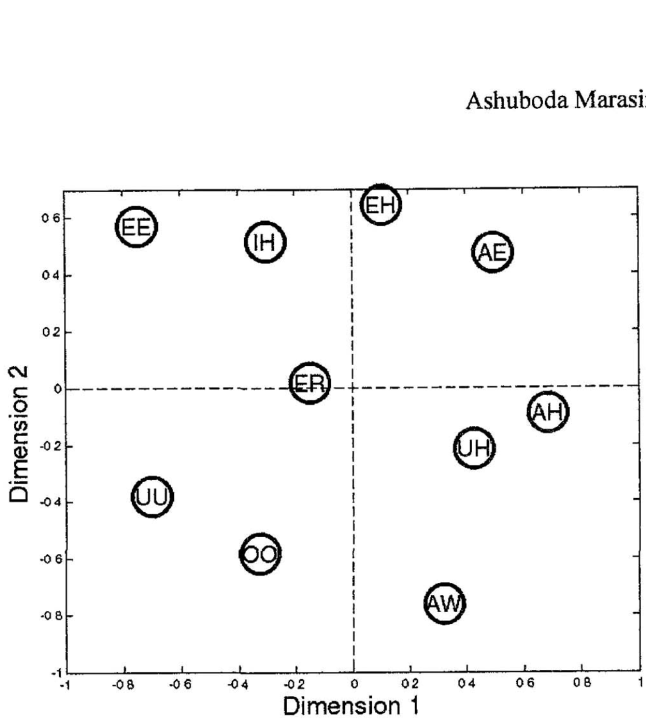


Figure 5. *INDSCAL*–derived perceptual space for native Japanese speakers.

Table 3

	UPS	English
English	0.9547	
Japanese	0.9185	0.8923

Correlation between dimensions of Universal Perceptual Space (*UPS*) and perceptual dimensions of English and Japanese languages.

This result, of course, does not prove that the most salient perceptual dimensions are the same for the two groups of subjects; rather the result provides no evidence that the groups differ in their global responses to the stimuli. The conclusion that the null hypothesis should be retained is not the same as proving that no differences exist.

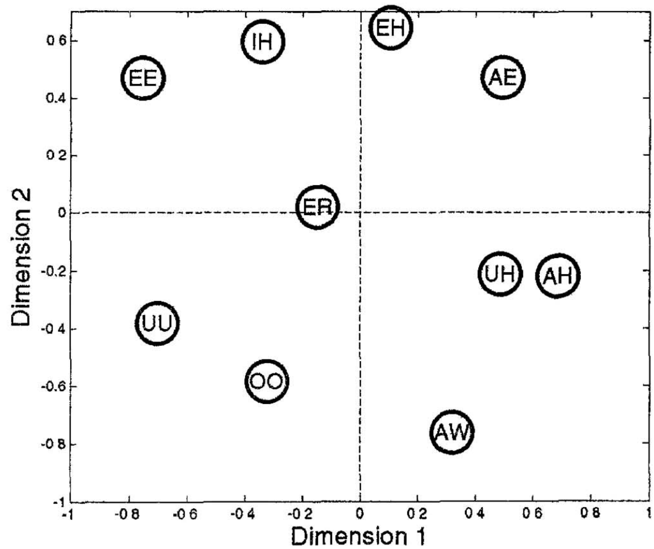


Figure 6. *INDSCAL*–derived universal perceptual space for two languages. Note that the axes have been given generic labels here, since no interpretation of the dimensions is possible on the basis of dissimilarity data alone.

3.2. Semantic Differential Rating Result

Ratings on the 12 bipolar adjective scales shown in Table 2 were collected for the 10 vowel sound stimuli from the 28 English and 52 Japanese listeners. A single 10×12 matrix of bipolar adjective ratings data was constructed for each listener, and these were combined into a single submission for Principal Components Analysis *PCA*, using the *FACTOR* routine found in the *SPSS* statistical analysis software. Among all 12 adjectives, only thickness ratings did not significantly discriminate between vowel sound stimuli.

Multivariate ANalysis Of VAriance (*MANOVA*) shows a significance value of (p) = 0.341 for thickness. Since the significance level was set at $p<0.05$, the adjective rating of thickness clearly did not reach significance. Therefore, in subsequent analysis, the thickness ratings were not used. Figures 7 and 8 shows the *PCA*–derived adjective weights (loadings) based upon semantic differential ratings from the two groups of listeners. In contrast to the *INDSCAL*–derived perceptual space, which was shown to be common to the two groups of subjects, the semantic space of adjectives revealed by PCA are somewhat in conflict.

The weights for selected bipolar adjective scales were analyzed through the *MANOVA* routine found in the *SPSS* statistical analysis software. The null hypothesis would be rejected if significant differences were found between the two groups in their use of the adjectives. *MANOVA* showed that both groups differed significantly in their use of all adjectives except Bright ($p=0.685$), Heavy ($p=0.329$) and Powerful ($p=0.784$). This implies that the use of the 3 adjectives Bright, Heavy, and Powerful did not differ between the American and Japanese subjects. Therefore, the overall semantic weights space for adjectives was found not to be common between the English and Japanese languages.

3.3. Implications to Language Typology

Attempts to characterize language typologies and thereby propose linguistic universals are necessarily limited by the practice of sampling. Because a full assessment of over 6,000 languages is impractical, typologists have often based their work on a subset of languages selected to encompass structural and geographical diversity. Such work has yielded noteworthy results (Ladefoged 1996), which have shaped the ways in which other linguists characterize their own data in terms of markedness. Liljencrant's (1972) dispersion theory claims that a number of typological trends in the phonetic composition of vowel inventories can be derived from the assumption that the phonetic implementation of vowel categories is maximally dispersed in the available perceptual space.

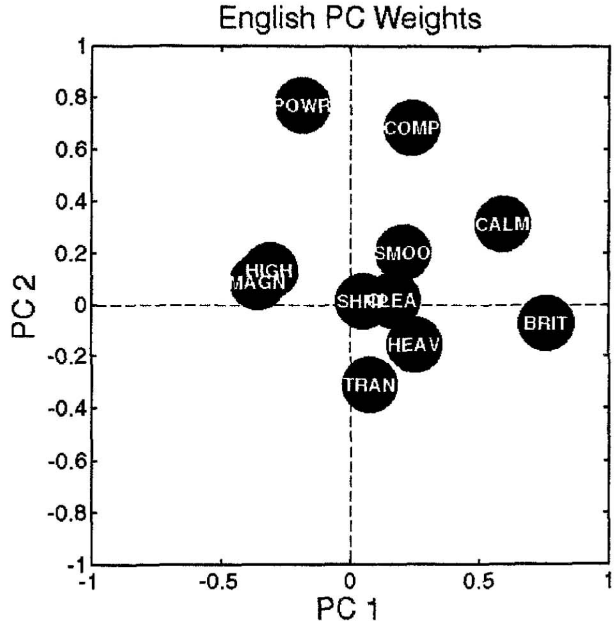


Figure 7. Principal–Component Weights for 11 bipolar adjective scales result for English native speakers. Thickness ratings were not used. PC1 and PC2 denote principal components 1 and 2.

For instance, if a language has three contrastive vowels, they are highly likely to be implemented as [i], [a], and [u], which represent the extremities of the F1-F2 triangle (F1 is first formant and F2 is second formant frequencies, Figure 11) in which vowels can be realized. The underlying idea is that dispersion leads to less confusion of contrastive vowels and thus to more effective communication.

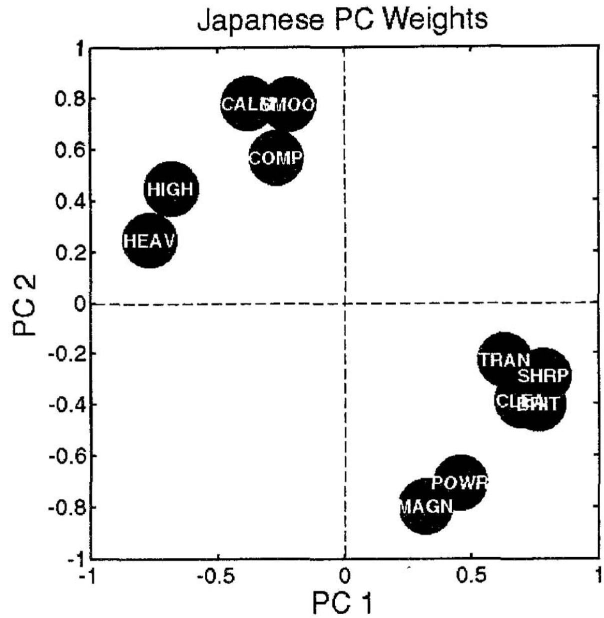


Figure 8. Principal-Component Weights for 11 bipolar adjective scales by Japanese native speakers. Thickness ratings were not used. PC1 and PC2 denote principal components 1 and 2.

One potential shortcoming of the sampling strategy stems from a question fundamental to linguistics: At what point do two similar speech varieties differ sufficiently such that they are deemed distinct? This matter is particularly relevant for phonological typology, as historically related varieties that exhibit syntactic, morphological, and lexical similarities are often classified as "same," despite differences in their sound systems. In such cases, the variety that typically earns representation in the literature is the standard language.

Vowel dispersion can emerge in a language without the need for assessment by individual speakers of the global properties of their inventories. The computational model on which this claim is based derives a number of additional predictions with varying degrees of accuracy about the typology of vowel inventories, considered in the following sections.

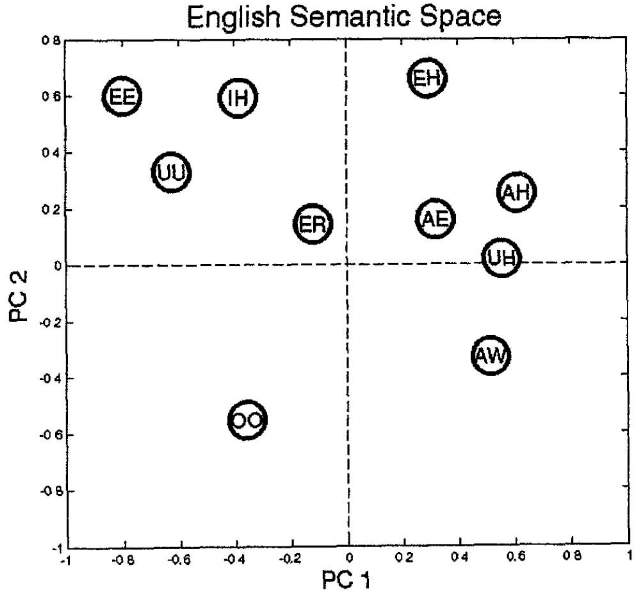


Figure 9. PCA–derived vowels semantic space for native speakers of American English. PC1 and PC2 denote principal components 1 and 2.

3.3.1. Typology in Human Vowel Systems

Humans are able to distinguish a huge number of different vowel sounds. According to Ladefoged and Maddieson (1996) there are languages that make five distinctions in the height of vowels, languages that make three distinctions in their position, and languages that make three distinctions in lip rounding. This would make for a total of at least 45 possible basic vowel qualities. However, any human language only uses a very limited subset of these. There are languages that utilize a large number of vowel phonemes (e.g., Punjabi, Hungarian, and Gaelic; Crystal 1997) but they use other processes—such as length, nasalisation, and pharyngelisation—besides quality to distinguish vowels.

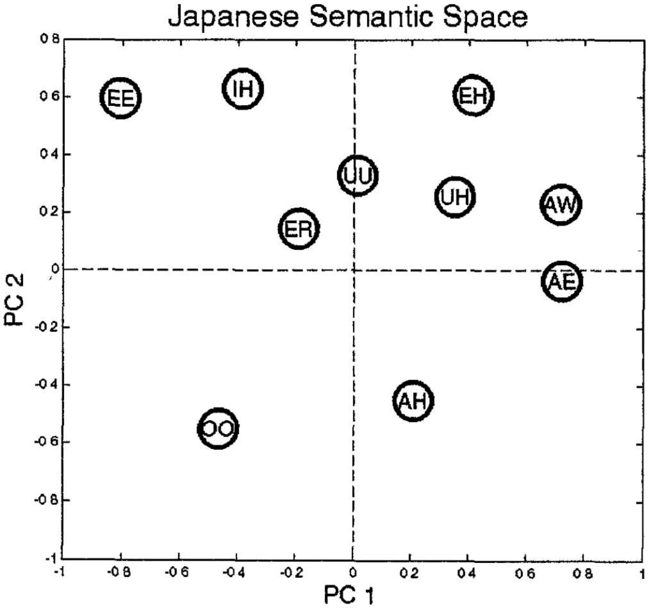


Figure 10. PCA–derived vowels semantic space for native speakers of Japanese. PC1 and PC2 denote principal components 1 and 2.

Furthermore, the small subsets of the possible vowels that languages use are not chosen at random (Crothers 1978). Some vowels appear more often than others, and vowel systems tend to be quite symmetrically well–crossed. For artificial languages, such as Unish or Esperanto, perhaps symmetric distribution of vowels and phonemes, in general, is appropriate.

Typologies of human vowel systems are based on phonetic descriptions of the vowel phonemes of languages. Phonemes are by definition minimal units of sound that can make a difference in meaning. However, it is quite possible that two speech sounds that are different (but close) phonetically do not make any distinction in meaning (De Boer 2001). These sounds are then called allophones of a phoneme. This happens for example through the influence exerted by neighboring sounds.

We investigated the semantic meaning of American English vowel sounds through our second experiment. Figures 9 & 10 show the *PCA*–derived semantic spaces for two languages, giving a useful visual description of the phonemes of a language. On the other hand, if one wants to classify languages based on which phonetic signals are used for realizing their vowel phonemes, a problem manifests. A choice needs to be made as to which phonetic realization is representative of the phoneme. Usually the most frequent allophone of a phoneme is taken to be the representative one. These representative allophones can then serve as a basis for a typology of possible vowel systems. Some researchers have even considered vowel systems with phonetically different elements as belonging to the same category (e.g., Crothers 1978, who characterizes [i], [a], [u] and [i], [a], [o] as belonging to the same type). It will be assumed here that this is a valid methodology. However, it should be kept in mind that a typology and classification of vowel systems is based ultimately on abstract phonemes. The actual observed signals in a language can be quite a bit messier than would be expected from a typological classification of the language.

A case in point is the vowel system of English. Figure 11 shows the formant space of 10 vowel sounds (Peterson 1952). This seems like a reasonably symmetrical ten vowel system. But if one looks at the reality of the figure, one sees that the actual formants (thinking about clusters, which are based on data from many different speakers) span quite a considerable area of the acoustic space, meaning that the vowels could have been labeled differently as well. Also there might to be overlap between the different vowel clusters, signifying that it is not always possible to say to which phoneme a given signal would have to be mapped (but probably this overlap disappears if higher formants are also taken into account).

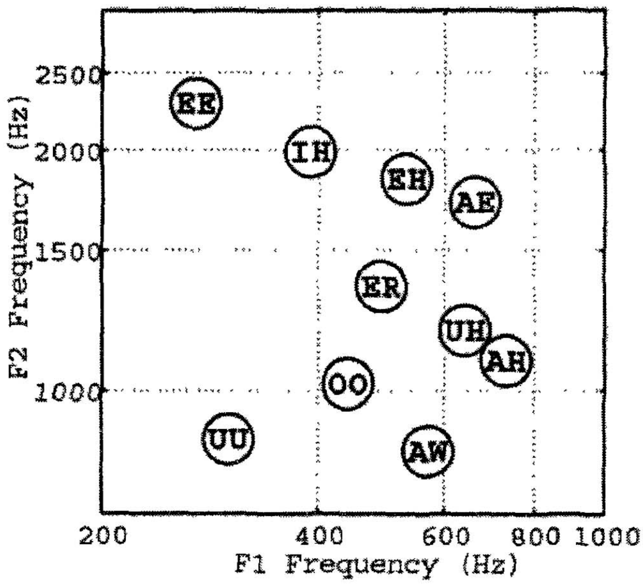


Figure 11. Formant space for 10 American English Vowels (Peterson & Barney 1952).

3.3.2. Vowels Typological Classification Based on the Observations

The basic principle of vowels is contrast: two sounds are in contrast if they can occur in the same environment and are not generally in free variation. This principle, modified to allow for partial overlap of phonemes, is probably sufficient in the majority of cases to decide whether the difference between two sounds is phonemic or not (Chomsky 1968). There are both practical and theoretical reasons for basing phonological typology on the classical phoneme (De Boer

2001). On the practical level, the majority of modern language descriptions have a phonemic basis. On the theoretical level, phonemic analysis answers a basic phonological question: what phonetic features are employed distinctively in a language? The abstract systematic phonemes of generative phonology do not provide an answer to this question since they do not express phonetic features in direct way. In a search for answering this question, we examined the phonological typologies through perceptual and semantic similarities of the vowel sounds.

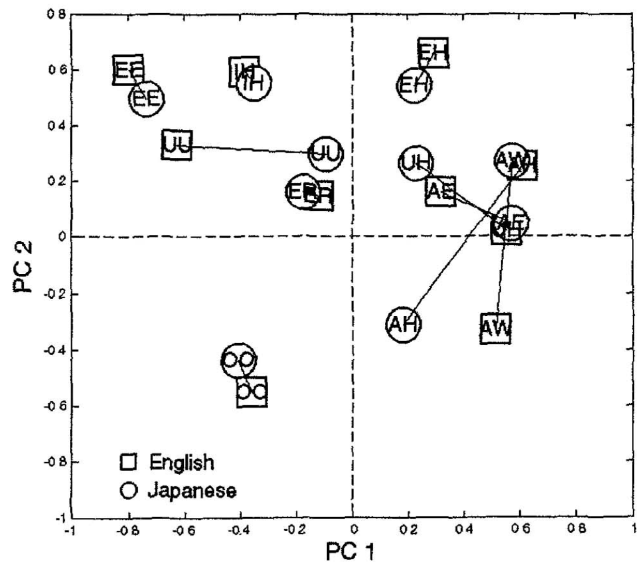


Figure 12. Common vowel semantic space of the 10 American English vowels as judged by native speakers of English (in squares) and Japanese (in circles).

Figure 12 shows the common vowel semantic space of the ten American English vowels as judged by native speakers of English and Japanese. These results reveal that, from the ten American English vowels, a shared vowel semantic space of the two language groups was realized for the five vowels /EE/, /IH/, /EH/, /OO/ and /ER/ (see Figure 13). Examining the distinct phonetic features of these five vowels may help explain why they share a common semantic space. Although the articulatory features of these five vowels with regard to lip rounding and tongue height do not reveal any marked synchronization (see Table 1), their formant structures reveal that they all are somewhat equally dispersed throughout the vowel space, particularly in the high-front through high-back regions. It may be that the auditory distinctness of these five vowels

Table 4

Vowel		Japanese	American
IPA symbol	ASCII code	%	%
i	EE	86	97
ɪ	IH	92	99
ɛ	EH	70	99
æ	AE	48	99
ɑ	AH	35	88
ʌ	UH	49	98
ɔ	AW	44	94
ʊ	OO	76	89
u	UU	69	86
ɚ	ER	59	99

The mean percentage of correct responses for 10 American English vowels as identified by a native Japanese and native American English listeners within a 10 AFC identification task (from Lambacher et. al. 2000).

makes it easier for Japanese listeners to be more attuned to the phonetic and acoustic representations of them. For example, the unusually high second formant of /EE/ and the unique character associated with the low third formant of /ER/ (Ladefoged 2001) may have resulted in the Japanese listeners judging them to be very similar semantically as that of the native English listeners.

Prior research by Lambacher, Martens & Molholt (2000) may also provide further explanation for the shared common semantic space of the five vowels /EE/, /IH/, /EH/, /OO/, and /ER/ as judged by listeners of the two language groups. In their cross-linguistic study, the performance of native Japanese and native American English listeners was examined in identifying the same ten American English vowels of the present study. Table 4 shows the mean percentage of correct responses for ten American English vowels as identified by the Japanese and American English listeners within a ten-alternative forced-choice (10 AFC) identification task. Notice that of the ten vowels tested, the Japanese listeners identified four out of the five vowels /EE/ (86%), /IH/ (92%), /EH/ (70%), and /OO/ (76%), which were judged in the present study as semantically similar by the listener groups, more accurately than the other English vowels, except /ER/ (59%). Superior identification performance by two or more native listener groups for a particular set of vowels presumes that their vowel qualities are more phonetically salient, and it perhaps may lend credence to the theory of a phonological and semantic typology in human vowel systems.

The perceptual model of vowel systems presented above relates phonological structure to factors operating in the ordinary use of language. Stated simply, the idea is that since the linguistic function of sounds is to distinguish different meaningful elements, one would expect the dominant types of phonological system to be those which make the most efficient use of the human sound production and perception abilities. In contrast, generative phonologists have sought to reduce phonetic description to feature systems (Crothers 1978, De Boer 2001), with a relatively limited number of features and distinctions within each feature. While properly defined features have the virtue of characterizing the kind of contrasts or types of rules that typically occur in languages, the justification for a particular system of features and the system of relations between features is not readily apparent in the features themselves. For example, the features—

high, low, back, round—interpreted as dichotomized features, are obviously largely sufficient for characterizing the kind of vowel contrasts that occur in the world's languages. Similarly, conventional notation fails to account for the observation that front vowels are less complex than mid-back vowels.

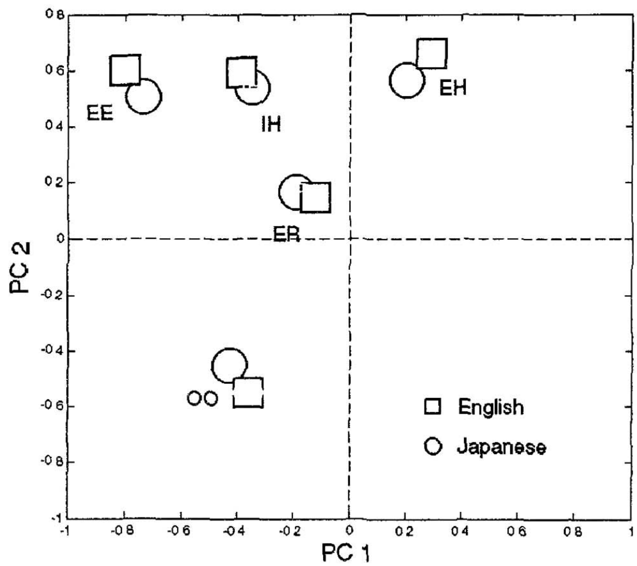


Figure 13. The five most commonly perceived English vowels, as judged by native speakers of English (in squares) and Japanese (in circles), with regard to vowel quality (i.e., semantic space).

Universal tendencies of human perception introduce some well-established generalizations about the phonological typology of vowel systems. Examining the perceptual space of American English as judged by different listener groups is most suitable because the English language contains a relatively large vowel inventory that represents a wide-ranging vowel space in comparison to other languages of the world. A universal perceptual space provides a strong claim of phonological typology similarities regarding to the vowel inventory of American English as perceived by native speakers of English and Japanese. The *INDSCAL*-derived universal perceptual space shows the typological comparison between two listener groups in their perception of the vowel space of American English vowels. From the *INDSCAL* analysis, statistical generalizations were found about the typology of vowel inventories. This might help to draw an analogy between language typologies by mapping the multidimensional vowel perceptual space. Of course, further research is required to test these implications on other sets of vowels with other listener groups. In addition, semantic differences were found between listener groups in their use of adjectives to describe the target American English vowels, although there were similarities between both groups in their use of 3 of the adjectives. Further research is required to help clarify the reasons why these 3 adjectives were used in a similar way.

4. Conclusion

Taken together, the results of the *INDSCAL* and the *PCA* analyses support the conclusion that native speakers of English and Japanese use adjectives differently in describing the same perceptual dimensions. In addition, ratings on 12 bipolar adjective scales for the same set of sounds showed that the English and Japanese semantic scales related somewhat differently to the dimensions of their shared perceptual space, even though a few adjectives—such as *Bright, Heavy* and *Powerful*—were used to describe similar perceptual variations. It was not assumed at the outset that ratings on these 12 adjective scales should necessarily capture the most salient differences between the stimuli. Regarding the potential for generalizing semantic differential ratings obtained in one language to aid in the interpretation of data from listeners speaking a different native language, the results of this study suggest that caution be exercised. It was hoped that the results would provide a better understanding of how bipolar adjectives from each language are used by listeners to describe the sound of American English Vowels. In addition, a single, universal perceptual space for American English vowel sounds was derived for two listener groups, who were judged to share similar global perceptual responses due to the overlap of the individual dimensional weights obtained from the *INDSCAL* analysis. Similarly, a shared vowel semantic space of the two language groups was realized for the five vowels /IH/, /EE/, /EH/, /OO/ and /ER/. Taken together, the results of the two experiments illuminate the typological implications of a common perceptual space among a set of vowels in a given language as judged by two listener groups. They also provide some insight into typological facts that should be taken into account in a theory of phonological typology.

References

Borg, I. & P. Groenen. 1997. *Modern Multidimensional Scaling: Theory and Applications*. New York: Springer-Verlag.

Chomsky, N. & M. Halle. 1968. *The Sound Pattern of English*. Cambridge, MA: MIT Press.

Comrie, B. 1981. *Language Universals and Linguistic Typology*. Chicago, IL: University of Chicago Press.

Crystal, D. 1997. *The Cambridge Encyclopedia of Language*. (2nd edition). Cambridge: Cambridge University Press.

Crothers, J. 1978. Typology and Universals of Vowel systems. In J. Greenberg, C. Ferguson, & E. Moravcsik (eds.), *Universals of Human Language* 93-152. Stanford, CA: Stanford University Press.

De Boer, B. 2001. *The Origins of Vowel Systems*. Cambridge: Oxford University Press.

Keweley-Port, D. & B. Atal. 1989. Perceptual Differences between Vowels Located in a Limited Phonetic Space. *Journal of Acoustic Society of America* 1726-1740.

Kister, D. & F. Wightman. 1991. A Model of Head-related Transfer Functions Based on Principal Components Analysis. *Journal of Acoustic Society of America* 1637-1647.

Ladefoged, P. & I. Maddieson. 1996. *Sounds of the World's Languages.* Oxford: Blackwell Publishers.

Ladefoged, P. 2001. *Vowels and Consonants.* Oxford: Blackwell Publishers.

Lambacher, S., Martens, W. & Molholt, G. 2000. A Comparison of Identification of American English Vowels by Native Speakers of Japanese and English. *Proceedings of the Phonetic Society of Japan Meeting* 213-218.

Liljencrants, J. & B. Lindblom. 1972. Numerical Simulation of Vowel Quality Systems: The Role of Perceptual Contrast. *Language* 48, 839-862.

Martens, W. 1987. Principal Components Analysis and Resynthesis of Spectral Cues to Perceived Direction. Presented at International Computer Music Conference, Urbana-Champaign, IL.

Martens, W., C. Marasinghe., C. Giragama., & A. Madurapperuma. 2000. Topic-dependent Adjective Use in Japanese and Sinhala: Selection of Adjectives Differentiating Guitar Sounds. *Seventh International Workshop on Human Interface Technology* 27-34.

Martens, W. & A. Marui. 2002a. Multidimensional Perceptual Scaling of Tone Color Variation in Three Modeled Guitar Amplifiers. *Proceeding of 112th Convention Of the Audio Engineering Society* (Preprint) 5552.

Martens, W. & C. Giragama. 2002b. Relating Multidimensional Semantic Scales to a Common Timbre Space. *Proceedings of 113th Convention of the Audio Engineering Society* (Preprint) 5705.

Osgood, C., G. Suci., & P. Tannenbaum. 1957. *The Measurement of Meaning.* Urbana-Champaign, IL: University of Illinois.

Osgood, C., W. May., & M. Miron. 1975. *Cross-Cultural Universals of Affective Meaning.* Urbana-Champaign, IL: University of Illinois.

Peterson, G. & H. Barney. 1952. Control Methods Used in a Study of the Vowels. *Journal of Acoustic Society of America* 175-184.

Ramsay, J. 1980. The Joint Analysis of Direct Ratings, Pairwise Preferences, and Dissimilarities. *Psychometrika* 149-165.

Sandell, G., & W. Martens. 1995. Perceptual Evaluation of Principal-Components Based Synthesis of Musical Timbres. *Journal of Audio Engineering of America* 43, 1013-1028.

Shepard, R., A. Romney, & S. Nerlove. 1972. *Individual Differences and*

Multidimensional Scaling. New York: Seminar Press.
Shiffman, S., M. Reynolds., & F. Young. 1981. *Introduction of Multidimensional Scaling*. New York: Academic Press.
Solomon, L. N. 1958. Semantic Approach to the Perception of Complex Sounds. *Journal of Acoustic Society of America* 5-30.
Tucker, L. & S. Messick. 1963. An Individual Differences Model for Multidimensional Scaling. *Phychometrika* 38, 333-368.