Language Universals: Cross-lingual Comparison of Topic Dependent Adjectives

Charith Giragama[1], William Martens[2], Dishna Wanasinghe[1], Michael Cohen[1] & Chandrajith Marasinghe[1]

University of Aizu[1], McGill University[2]

Abstract

This study examined the multilingual subjects' perceptual similarity, and semantic relation in order to find universals of languages. Two Indo-European languages (English and Bengali) and one non-Indo-European language (Japanese) were used. A common perceptual space was obtained through Multidimensional Scaling (MDS) analysis method. In contrast, semantic approach was taken using Semantic Differential Analysis (SDA) method. Apart from these, a least square mapping method, Procrustes analysis, was also employed to determine whether semantic scales relate differently to the dimensions of the common perceptual space that did not differ between these three groups of languages. Breaking the tradition of taking vowel sounds to find relation between languages, however, a sophisticated approach was taken by having six guitar sounds as stimuli, which could be considered as a common parameter for all subjects of different groups of languages. Ten topic-dependent bipolar adjective scales on which responses are taken, were constructed in a preliminary study. The results of the SDA analysis showed that

the English, and Japanese semantic scales related differently to the MDS derived common perceptual space, while the Bengali semantic scales related somewhat similarly to the same common perceptual space. And also Procrustes Analysis revealed that the Bengali factor space is closer to the English factor space than to the Japanese factor space. With regard to 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 the current study suggest that caution be exercised. Moreover the results of this study suggest that interpreters will have to be cautiously attentive in interpretation of data from Indo-European family of languages (English and Bengali) to non Indo-European family of languages (Japanese).

Keywords: multilingual, automatic translation, semantic differential analysis, natural language processing, multidimensional scaling

1. Introduction

The novelty in the current study was to compare semantic differential scaling results between two Indo-European languages (English and Bengali) and a non-Indo-European language (Japanese). Though the number of recognized language families in the world today are many, more than half of the world's population speaks an Indo-European tongue as a first language (Rosen 2003). The languages studied here are, English, Bengali, and Japanese are among the top 10 languages in the world. About 450 million speak English language, about 200 million people speak Bengali, and nearly 125 million speak Japanese.

Cognition (psycholinguistic results of perception) and the relation between languages have been increasing in studies of cognitive linguistics over the last couple of decades (Edwards 1997, Wierzbicka 1999). How do people talk about what they see or hear and what they do choose to say about it? More specifically, how does

the mind perceive spatial information, how does it encode linguistic information, and finally how does it communicate between these two? In a psychological point of view, this central issue could be addressed by observing what goes into the organism (stimuli) and what comes out (responses). Between these two observation points lies an unknown, the nervous system. Nowadays it is fashionable to refer to this region as "a little black box," the same refer as "perception." Most psychological theories are made up of hunches about what goes on in this "little black box." But for deeper understanding of those perceptions, it is necessary to use a measurable response output from it. Although subjective response output is most naturally and directly assessed through the medium of language, in cross-lingual perceptual comparisons that which has been called the "language barrier" must somehow be pierced to get the most natural properties out of the mind of multilingual subjects. In the studies of multilingualism, cognitive research has focused mainly on semantic representation. This study focuses on semantic representation in order to find universals of languages.

1.1. Language Universals

Nearly five thousand languages are spoken in the world today. Although they seem to be quite different, still many of them show similar principles. However, language universals research is concerned with finding patterns that are common to all human languages. The approach to language universals pioneered by Joseph H. Greenberg (Greenberg 1993, 1966, 1978) and others (Comrie 1984, 1989; Croft 1990). Language universals are just one facet of linguistic theory.

The number of absolute universals is, however, relatively small; linguistic theories are usually built on these universals since they tend to be true of all languages. More commonly occurring universals, the ones, which are characteristic of most language universals research, are implicational. Implicational universals can only be es-

tablished through a comparison across a wide range of languages. Our aim in this study is to identify semantic categories that are shared and referred by three out of top ten languages in the world, English, Japanese and Bengali. By establishing conceptual connection between semantic categories that justify similar patterns among these three languages, ultimately it could be hoped to arrive at a detailed picture of conceptual distinctions and conceptual groupings that are important to humans in creating a new universal language that ends up being speakable.

2. Analysis Methods

The working assumption in the current study is that complex sounds having multiple 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 sounds 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.

In the current study both MDS based and SDA based methods were employed. Both of these methods have limitations, which are largely circumvented when they are combined (e.g., via joint analysis (Ramsay 1980) or via external unfolding Meulman et al. 1986). MDS analysis of dissimilarity ratings are often included in investigations of complex perceptual phenomena in order to indicate the involvement of stimulus parameters for which direct ratings might not be collected. Conversely, the dissimilarity-based perceptual

space can reveal which stimulus parameters do not enter into the listener's global evaluative reactions. Also, wide variation in ratings on a particular adjective scale might not correspond to large perceptual differences. Therefore, while adjective ratings aid in the interpretation of the MDS derived perceptual dimensions, the dissimilarity ratings aid in identifying which of the adjective ratings scales correspond to salient perceptual attributes. The ten adjective scales chosen for direct rating in the current study were based upon the results of a previous study (Martens & Giragama 2002, Giragama 2002), these scales being formed from frequently chosen adjectives. However, it was not assumed at the outset of the current research that ratings on the pre-selected adjective scales should necessarily capture the most salient differences between the stimuli. Rather, the goal was to explore which of the adjective scales would correlate most highly with the MDS derived perceptual dimensions, and thereby aid in the interpretation of the multidimensional perceptual space for the stimuli. Also, the relative salience of the MDS derived perceptual dimensions should serve to identify which of the adjective scales might correspond to the biggest differences between the stimuli. While the full range of a given adjective scale might be used in the case of all ten bipolar adjective pairs, this is no guarantee that the perceptual differences along each scale would be of equivalent psychological magnitude. However, via the MDS analysis of the obtained dissimilarity ratings, there is a basis for determining how perceptually distinguishable multidimensional stimuli are from each other with respect to the adjective scale values. For example, the MDS based spatial configuration for the stimuli can potentially reveal which stimulus parameter variations contribute most to the listener's global evaluative reactions. This is one of the four primary purposes of such dissimilarity analysis identified in the text on "Modern Multidimensional Scaling" by Borg & Groenen (1997).

The following three subsections briefly introduce the three multivariate analytical techniques employed in this research project.

2.1. MDS and INDSCAL

INdividual Differences SCALing (INDSCAL) is one of the more popular implementations of the scaling technique termed MDS, which is a generic term for a data analytic method that can be used to derive a spatial representation for a set of stimuli based upon a single set of measured similarities (or dissimilarities) between those stimuli. Four primary purposes of such perceptual scaling have been identified as follows (Borg & Groenen 1997):

- To create a low-dimensional representation of otherwise complex data,
- To test how distinguishable multidimensional stimuli are from each other,
- To discover the stimulus dimensions that underlying judgments of (dis)similarity,
- To model the psychological dissimilarity between stimuli in terms of a distance function.

Of course, individual subjects may differ in how they form judgments of global dissimilarity, and so a refined method for doing a weighted MDS analysis (Shiffman et al. 1981) that takes such individual differences into account is to be recommended. This paper teaches the use of INDSCAL (INdividual Differences SCALing) (Carroll & Chang 1970) analysis 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 which may place different weights upon each of the perceptual dimensions.

Interpreting the results of classical MDS is problematic because the solution can be rotated through an arbitrary angle without violating the structure of the solution. Of course, inter-stimulus distances remain invariant under rotation of both classical MDS and INDSCAL solutions

alike; but the orientation of the INDSCAL solution is determined by modelling agreement between subjects. INDSCAL is designed to separate those factors that are common to a group of subjects from the ways in which subjects differ. The mathematical basis for these advantages are well explained in the book by Borg & Groenen (1997), and are beyond the scope of this paper.

2.2. PCA (Principal Components Analysis)

PCA was used in this study to find the semantic components underlying adjective ratings made for a small set of stimuli for each of three groups of listeners, both separately and when pooled together into a single group. PCA is a common procedure in multivariate statistics (Harris 1985) that is a tool that systematically identifies redundancy and offers a method for reducing the dataset containing a large number of variables to a smaller number of components. Its application to describing sound quality of sound-reproducing systems is well established (e.g., see Gabrielsson & Sjogren 1979). Similar applications to multilingual semantic scaling can be found in the work of Namba et al. (e.g., Namba et al. 1991).

PCA as a statistical technique was first described in Hotelling (1933), who applied it to the scoring of intelligence tests. PCA's essential task is to reduce a matrix containing correlated column vectors to form a set of orthonormal basis functions using the eigenvectors of the covariance matrix of the input data. PCA yields two matrices of data, called scores (or basis vectors) and weights (or loadings); the principal component (PC) scores can be calculated via post-multiplication of the input data by the PC weights. For the current project, the PC scores for each stimulus are regarded as the coordinates for each stimulus in a Semantic Space. The PC weights, on the other hand, show the relationship between the adjectives for which ratings have been obtained, akin to the factor loadings of Factor Analysis.

2.3. MANOVA (Multiple Analysis of Variance)

Multiple ANalysis Of VAriance, or MANOVA, was used in this study to test the equality of mean vectors for two or more populations. For example, if for the three language groups are tested, only univariate data had been collected, then a simple ANOVA for the three cases could be run; however, in all comparisons made here, the groups were compared on multiple response variables simultaneously. The research questions addressed here were two: Are the dimensions of perceptual space for a small set of stimuli common among groups of listeners with differing native languages? Do the words used by multilingual groups of listeners share common underlying semantic structures when used to describe that same small set of stimuli? For both questions, the null hypothesis, that no differences exist, can be retained only if no evidence is found to reject the null, and accept the alternative hypothesis, that be retained is not the same as proving that differences do exist between the groups.

3. Experiment Method

Stimulus selection was based upon the results of a prior study (Martens et al. 2000b) of the sound of guitar effects in which untrained listeners were presented with stimuli submitted to either distortion-based or modulation-based effects processing. The stimulus selection included only six classic distortion effects (featured in the MIDI-controllable Boss GX-700 Guitar Effects Processor (Boss 1999)). Adjective selection was also based upon the results of a prior study (Martens et al. 2000a, Giragama 2002). In the first phase of this project, two adjective selection methods were compared. One was the method of selected description of Namba (Namba et al. 1991). This method found what adjectives were produced in response to the stimuli when no distinction between the stimuli was

required, in contrast to the triadic comparison method (Berg & Rumsey 1999) in which adjectives distinguishing between stimuli were to be chosen. This latter method provided a better means for generating adjectives for consideration as potential candidates for the anchor points of attribute rating scales. The adjectives generated in response to the question of how an odd stimulus differs from other stimuli are potentially more useful in differentiating between stimuli than adjectives, which are simply descriptive of a stimulus presented in a single trial, in isolation from other stimuli.

3.1. Dissimilarity Rating

Dissimilarity ratings for deriving Perceptual Space were collected for 77 listeners in three roughly equal-size groups: 26 native speakers of English, 26 native speakers of Japanese and 25 native speakers of Bengali. In order to do pair-wise comparisons, subjects were presented with 30 pairs of stimuli on which they were supposed to give interpretation of the terms "Similar" or "Dissimilar" without disturbing their natural way of thinking. Listeners were asked to give global dissimilarity ratings on a five-point scale for all pair-wise comparisons of the guitar sound stimuli. The instructions were to listen to each stimulus pair once, and then 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" or "indistinguishable," 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 over stereo loudspeakers, always with a 1 second inter-stimulus interval and a 5 second inter-trial interval.

3.2. Semantic Rating

The impressions of the same set of sounds used for dissimilarity

ratings were measured by a 10 bipolar adjectives, 5 steps semantic differential scale (for deriving Factor Space). Data was collected for 66 listeners (22 native speakers of Japanese, 22 native speakers of English and 22 native speakers of Bengali) from the three groups.

Table 1. Corresponding bipolar adjective pairs in three languages, listed in order of presentation for subsequent semantic differential ratings.

English	Japanese	Bengali
clear - unclear	sunda - nigotta	poriskar - oshaccha
clamorous - quiet	souzoushii - ochitsuita	shorgolpurno - shanto
diffuse - compact	hirogatta - kojinmarishita	brisrito - ghano
hard - soft	katai - yawarakai	shokto - norom
weak - strong	yowai – tsuyoi	durbol - probol
light - heavy	karui - omoi	halka - vari
sharp - dull	surudoi - nibui	tibro - nishprovo
pleasant - annoying	urusai - kokochiyoi	monorom - biroktikor
thick - thin	usui - atsui	mota - patla
rough - smooth	aria - namerakana	omosrin - mosrin

Listeners were asked to rate each stimulus on the ten bipolar adjective scales (see Table 1 for the bipolar adjective scales). A 5 steps semantic differential scale was employed, a response of "1" indicated that the 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. Listeners were instructed to give a response of "3" if neither of the anchoring adjectives characterized the stimulus. The six stimuli were presented in a random order for each of ten adjective pairs, with a 5-second of interval between each individual stimulus presentation.

4. Perceptual Space

Separately derived Perceptual Spaces for each of the languages through INDSCAL are illustrated in Figure 1. Since all three spatial representation show similar shapes it's possible to create a common space having all three languages together. Thus, the Common Perceptual Space (Figure 1, above left) is derived by examining the ways in which subjects had made their judgments on dissimilarity for pair-wise comparison of stimuli, for determining whether three groups of listeners share single common perceptual space. From the dissimilarity data, a separate 6 x 6 (6 stimuli x 6 stimuli) matrix was constructed for each of 77 subjects. For these 77 participants, the 6 x 6 matrices were combined into a single 462 x 6 matrix, and this 462 x 6 matrix was submitted for INDSCAL analysis. The results of common perceptual space and subject space provide no evidence that the three groups differ in their global response to the stimuli. Note that the coordinates of all three Perceptual Spaces were highly correlated, ranging from $r = 0.88$ to $r = 0.99$ (see Table 2). The procrustean based transformation of perceptual space of each language separately to match with the perceptual space derived all languages

Table 2. Observed correlations between dimensions of Common Perceptual Space (CP Space) and perceptual dimensions of three languages

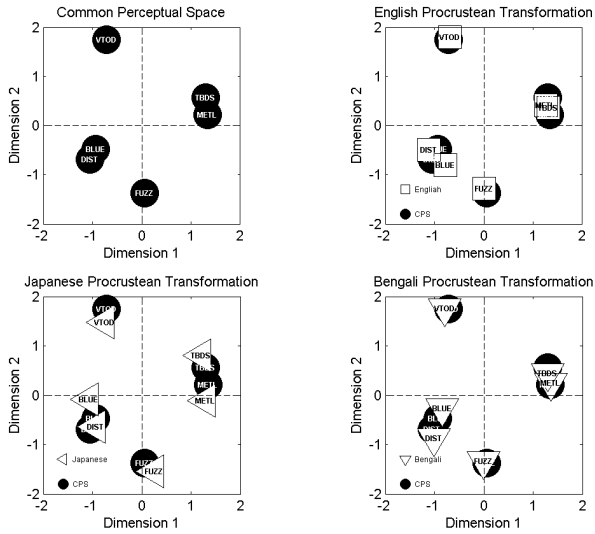|	CT Space	English	Japanese
English	0.98		
Japanese	0.88	0.90	
Bengali	0.99	0.97	0.91

Figure 1. The "Perceptual Space" common to the three groups of subjects, derived via (INDSCAL) analysis of dissimilarity data from 77 subjects and the result of Procrustes-based transformation of Perceptual Space computed from English, Japanese and Bengali to match the Common Perceptual Space computed from all languages together. CPS stands for common perceptual space. Codes are represented as follows "VTOD" for Overdrive, "BLUS" for Blues, "DIST" for Distortion, "METL" for Metal, "FUZZ" for Fuzz and "TBDS" for Turbo Distortion.
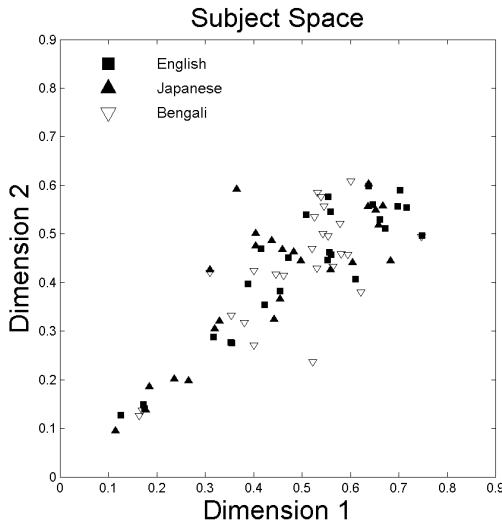
Figure 2. Subject Space for 77 subjects whose dissimilarity ratings were submitted to a single INDSCAL analysis. Weights on the two dimensions are plotted using separate symbols for each language.

together are depicted in Figure 1. The results showed that the coordinates of each spatial configuration match very well.

Subject Space derived using INDSCAL based upon the dissimilarity judgments is shown in Figure 2. Plotted here are the weights placed by each subject on each dimension in generating their dissimilarity ratings. The results of common perceptual space and subject space provide no evidence that the three groups differ in their global response to the stimuli.

Using MANOVA to test for differences between INDSCAL derived weights for individuals from different language groups, a further confirmation of a Common Perceptual Space for the three groups was observed. If individuals from the three language groups had produced dissimilarity based upon different perceptual weights,

then there should be greater variance between groups than within groups. Here, the null hypothesis was that the judgments from the three groups came from a single population, rather than three. The result, $[F(2, 15) = 0.0017, p(\text{Type II error}) < .05]$, provided no evidence for rejecting the null hypothesis. Thus it was concluded that the three groups probably do not differ in terms of their perception of the six stimuli presented.

5. Factor Space

The results of Factor loadings (Figure 3) revealed some agreement between English and Bengali language adjective use in two cases: both groups showed high loading on *sharpness, weakness, clamorousness* and *roughness*, the results of correlations of Factor loadings are tabulated in Table 3 was showed that Factor 1 of English was correlated at r = 0.78 with Factor 1 of Bengali. Similarly, Factor 2 of English was correlated at r = 0.71 with Factor 1 of Bengali. Although, neither Factor 1 nor Factor 2 of Japanese shows any correlation with Bengali, the correlation between Factor 2 of Japanese and Factor 2 of English was higher at r = 0.73. It is important to consider that semantic representation clearly varied between languages that are not closely related.

The Bengali group received significant loading on adjective scales that translate into English as *pleasant, weak* and *sharp,* English and Japanese groups received significant loading on English adjective scales *sharp* and *weak* (highest loadings for each principal component for all languages are shown in Table 4). The *diffuse* scale loaded almost neutrally for Bengali, but not for groups of Japanese and English speakers. The ratings on the *clamorous* scale received higher loading on Factor 2 for English, Japanese and Bengali (see Figure 3).
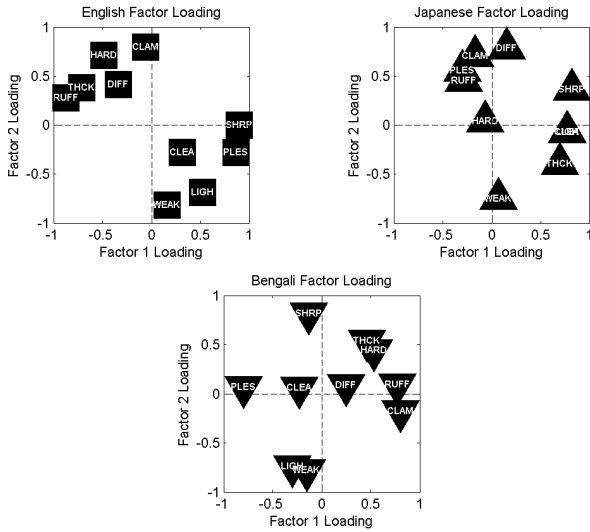
Figure 3. Factor 2 loadings were plotted against Factor 1 loading for ten bi-polar adjective scales calculated separately for each of the three groups of the languages. Codes are represented as follows "CLEA" for clearness, "CLAM" for clamorousness, "DIFF" for diffuseness, "HARD" for hardness, "WEAK" for weakness, "LIGH" for lightness, "SHRP" for sharpness, "PLES" for pleasantness, "THCK" for thickness and "RUFF" for roughness.

The loadings on Factors 1 and 2 are plotted for all three groups in Figure 3. In Japanese, sounds that were described as *rough* could also be described as *pleasant*, but not as *thick.* In contrast, sounds that were described as *rough* by English speakers were also *thick* but not *pleasant.* Furthermore, sounds that were *strong* might also be interpreted as *clamorous* by Japanese speakers, but not as *sharp.* In Bengali, sounds those were *strong,* were also *sharp,* but not particularly *clamorous.*

The relationship between Semantic Space and the INDSCAL de-rived Common Perceptual Space was examined by looking at their correlations between Perceptual Space coordinates and Factor scores within each group of listeners. In this way, the strongest semantic factor was identified as *sharpness* for Japanese Speakers and the correlation of this *sharpness* with dimension 1 of Perceptual Space was r = -0.84. The second strongest semantic factor was *diffuseness* and it was correlated at r = -0.75 with dimension 1 of the Perceptual Space. For Bengali Speakers, *sharpness* was identified as the strongest semantic factor and it shows r = -0.74 correlation with di-mension 1 of the Perceptual Space. *Diffuseness* was the second highest semantic factor and it was correlated at r = -0.53 with IND-SCAL dimension 2 of perceptual Space. No semantic factor showed a particularly high correlation with either dimension 1 or dimension 2 for English native speakers: that for *clamorous* was the highest with a correlation of r = 0.46 with dimension 2 of the Common Per-ceptual Space. The second strongest correlation was found between *hardness* and dimension 1 of the Perceptual Space at r = 0.40.

Further comparison of semantic spatial configurations was done using the MATLAB (Math 2002) "Procrustes" routine. Procrustes Analysis (least-squares orthogonal mapping) is a method used to find matching configurations of points between two data sets by ro-tation, translation, and scaling of one data set to match the other.

Table 3. Observed correlations between Components of Principal Components (Weights) derived via Principal Component Analysis (PCA) for three languages

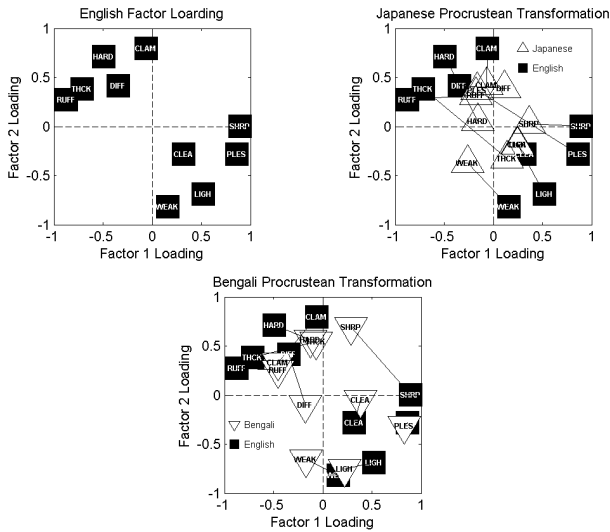| Factor Loading		Japanese		Bengali	
		Factor 1	Factor 2	Factor 1	Factor 2
English	Factor 1	0.30	0.05	**0.78**	0.37
	Factor 2	0.43	**0.73**	**0.71**	0.49
Japanese	Factor 1			0.23	0.12
	Factor 2			0.11	0.28

Figure 4. The results of Procrustes-based transformation of "Factor Space" of three languages derived via PCA of semantic ratings from 66 subjects. Japanese and Bengali adjectives to match the Factor space computed from the adjectives of English.

Table 4. The English translation of adjectives receiving the three highest Factor loadings on Factors 1 and 2 for each of the three languages. (See Table 1 for the particular translations of these English language adjectives into the other three languages)

English		Japanese		Bengali	
PC1	PC2	PC1	PC2	PC1	PC2
sharpness	weakness	sharpness	diffuseness	clamorousness	sharpness
roughness	clamorousness	lightness	weakness	pleasantness	weakness
pleasantness	hardness	clearness	clamorousness	roughness	lightness

The procrustean based transformations of Factor Spaces computed from Japanese and Bengali adjectives to match the Factor Space computed from English adjectives are depicted in Figure 4.

6. Semantic Space

In contrast to the reported global differences between stimuli captured as dissimilarity judgments, adjective ratings were collected to reflect differences in the semantic structure underlying description of the sound stimuli in each of three languages. Though the 10 pairs of anchoring adjectives for the semantic scales in each of the three languages were judged to be direct translations of the English adjectives by bilingual advisors, the resulting 6 x 10 matrix of ratings for each subject showed a different pattern of results within each group. The PCA derived adjective scores, based upon semantic differential data sets from the three groups, each analysed separately, are shown in Figure 5.

All correlations between dimensions of semantic spaces of each three languages are shown in Table 5. Of those, Factor 2 of Bengali and Factor 2 of English show the highest correlation at r = 0.95. And also the Factor 1 of Bengali with Factor 1 of English show high correlation r = 0.75. Moreover, both Factors of English also showed a fairly high correlation with Factor 1 and Factor 2 of Japanese.

Even though the INDSCAL derived Perceptual Space (Figure 1) was seemed common to the all three groups of subjects, the semantic spaces derived by separate (PCA) for each group were rather different from one another.

Two sets of predictor variables for spatial configuration of the six stimuli are provided. One set is the coordinate of the six stimuli on two semantic factors, derived through PCA analysis. The other set is the coordinate of the six stimuli on two dimensions, derived through INDSCAL analysis. However, the special configurations of

the six stimuli with these two sets of two-dimensional coordinates are somewhat different.

In order to demonstrate the dissimilarity of these two special configurations, a linear transformation of PCA points was done over INDSCAL points, using the Procrustes routine. Hence it is concluded that the PCA derived Semantic Space for each language and
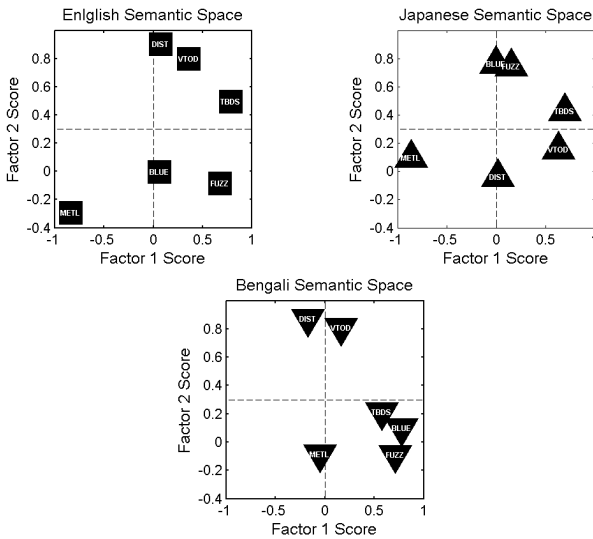

Figure 5. Semantic Space for each of three languages separately.

Table 5. Observed correlations between dimensions of Semantic Space for three languages

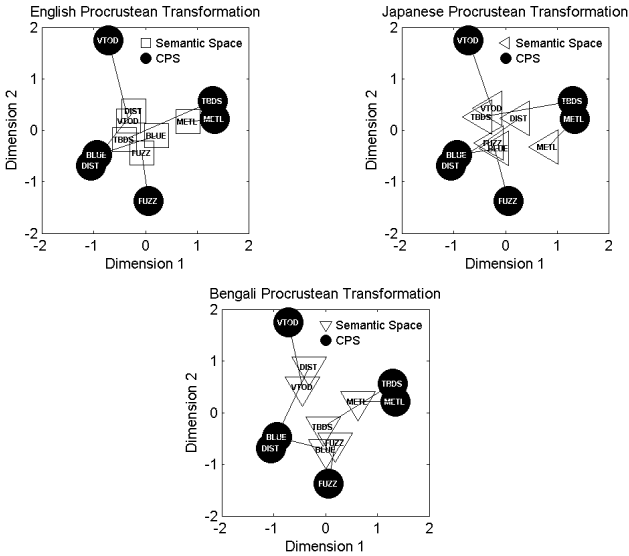| Factor Scores		Japanese		Bengali	
		Factor 1	Factor 2	Factor 1	Factor 2
English	Factor 1	**0.60**	0.46	**0.75**	0.17
	Factor 2	0.45	**0.82**	0.50	**0.95**
Japanese	Factor 1			0.36	0.45
	Factor 2			**0.78**	0.47

Figure 6. The result of Procrustes-based transformation of Semantic Spaces computed from English, Japanese and Bengali to match the Common Perceptual Space computed from all languages together.

INDSCAL derived Common Perceptual Space relate differently (see Figure 6). The translational component of the transformation was rendered negligible by standardizing the Factor score values.
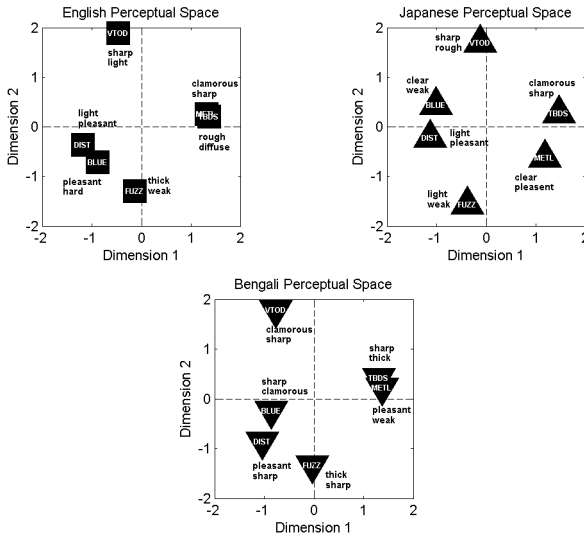
Figure 7. Perceptual dimensions for the sound effects derived from INDSCAL analysis of dissimilarity judgments obtained from native speakers of three languages. The adjectives written above of each sound effect symbol were those providing the clearest distinctions between each effect and the other effects.

7. Implications towards an Artificial Language

In the 17th century a group of mystics dreamt of universal concord. Among them Francis Bacon, René Descartes, and Gottfried Wilhelm Leibniz, proposed the construction of a so-called philosophical language that would consist of a system of communication based on classification according to logic rather than having arbitrary word creations as evolve in most natural languages. They wanted to propose that all mankind adopt a new science through a

perfect language that is being free from racial, national, religious and other prejudices.

A natural language would be one that evolves naturally as a more-or-less direct consequence of a communities need for internal communication with each other. An artificial language would be one that resulted from the deliberate actions of a person or a group of people. Earth is now a country with too many languages and it cannot expect to be united unless a universal language is found which tend to reduce the misunderstandings and antagonisms caused by language differences. A universal language could serve mankind as a "bridge language" to enable inter-communication between the numerous peoples of this planet who speak dissimilar languages. Not that an international language would be sufficient to turn the world into a utopia, but that it would be necessary to make it a tolerable place in which to live. A word in one language rarely means exactly the same thing as its closest counterpart in a different language. As a result of centuries of evolution, the "semantic space" of each word in a natural language is arbitrary. Thus, each word has built-in irregularities.

Universal languages that have had more popular success is the kind formed from elements or modified elements of existing natural languages. Esperanto, as an example, was deliberately designed as a step towards a religion of universal brotherhood. It is designed as an auxiliary language whereas its semantic space of each word is precisely defined in terms of the much more basic meanings of the components that make up each word.

Of course, many invented universal languages are devised with a view toward their being as natural as possible. But since many of them have their vocabulary and grammar based on those of the Indo-European tongues, speakers of non-Indo-European idioms find them difficult and even distasteful. The aim of the present study is to satisfy the obvious need of discovering language universals between Indo-European and non-Indo-European languages, which would better explain required details in synthesizing a new universal language that devised to be neutral for these two groups of speakers.

As visualized by the results of PCA in Figure 7, it is evident that the adjective *sharp* was found to usefully differentiate between sound Vintage Overdrive (VTOD); similarly the adjective *pleasant* was found well differentiate between sound distortion (DIST) by all three language groups, English, Japanese and Bengali. Moreover, it is noted that the adjective *thick* was used in a similar manner, in distinguishing the sound *fuzz* by native speakers of English and Bengali. Relying on the results of procrustean based transformations (Figure 4) and, the correlations between dimensions of semantic spaces of three languages (Table 5), it is tempting to conclude that the Japanese is semantically more closer to English than to Bengali. Nonetheless, the Bengali is semantically somewhat similar to English than to Japanese. Accordingly, the adjectives use in English language in distinguishing sounds has properties in common to both languages Japanese, and Bengali. All three language groups received higher loadings on the adjective scales that translate into English as *sharp*, *weak*, and *clamorous* as depicted in Figure 3. It can also be concluded that the group of stimuli might be well distinguished with the adjectives *sharp*, *weak*, and *clamorous* by English, Japanese, and Bengali speakers. The Language universals that have been proposed in our present study thus a suggestion that would provide a fruitful result in interaction between future language researches evolve in synthesizing universal languages.

8. Conclusion

The results of this study provide a database that might useful for research on language universals and is especially valuable in that the database is informed in a topic dependent manner. Three groups of listeners seemed to share similar perceptual responses to the stimuli, as revealed by the complete overlap of individual dimensional weights derived using INDSCAL. On the other hand, it was evident that semantic responses differed for each language, and different

languages express spatial relations in different ways even though they share a common similar perception. Factor scores from PCA on adjective ratings related differently to stimulus coordinates on the dimensions of a common perceptual space, except for the Bengali semantic spaces, which showed a quite similar relationship to the dimensions of Common Perceptual Space. With regard to 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. Moreover the results of this study suggest that interpreters will have to be cautiously attentive in interpretation of data from Indo-European family of languages (English and Bengali) to non Indo-European family of languages (Japanese). Furthermore, procrustean transformation showed that the Bengali semantically relate to English more than to Japanese.

Acknowledgement

References

Berg, J. & F. Rumsey. 1999. Spatial Attribute Identification and Scaling by Repertory Grid Technique and Other Methods. Presented at the 16th International Conference Of Audio Engineering Society. Rovaniemi, Finland.

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

Carroll, J. & J. Chang. 1970. Analysis of Individual Differences in Multi-dimensional Scaling via an N-way Generalization of "Eckart-Young" Decomposition. *Psychometrika* 35, 283-319.

Comrie, B. 1984. *Form and Function in Explaining Language Universals: In Explanations for Language Universals*. Edited by B. Butterworth, B. Comrie, & O. Dahl. Berlin: Mouton.

_____. 1989. *Language Universals and Linguistic Typology: Syntax and Morphology*. Oxford: Blackwell.

Croft, W. 1990. *Typology and Universals*. Cambridge: Cambridge University Press.

Boss Corp. 1999. *GX-700 Guitar Effects Processor Owner's Manual. Roland Corporation USA*, 5100 S. Eastern Avenue Los Angeles.

Edwards, D. 1997. *Discourse and Cognition*. London: Sage.

Gabrielsson, A. & H. Sjogren. 1979. Perceived Sound Quality of Sound-Reproducing Systems. *Journal of the Acoustical Society of America* 65, 1019-1033.

Giragama, C. 2002. *Relating Japanese and Sinhalese Semantic Scales to a Common Perceptual Space*. Master's Thesis. Aizu: University of Aizu.

Greenberg, J. 1963. Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements. In J. Greenberg (ed.), *Universals of Language* 73-113. Cambridge, MA: MIT Press.

_____. 1966. Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements. In J. Greenberg (ed.), *Universals of Language* 73-113. Cambridge, MA: MIT Press.

Greenberg, J., C. Ferguson, & E. Moravcsik. (eds.) 1978. *Universals of Human Language*. Stanford, CA: Stanford University Press.

Harris, R. 1985. *A Primer of Multivariate Statistics*. New York: Academic Press.

Hawkins, J. 1983. *Word Order Universals*. New York: Academic Press.

Hotelling, H. 1933. Analysis of a Complex of Statistical Variables into Principal Components. *Journal of Educational Psychology* 24, 417-520.

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

Martens, W., C. Marasinghe, C. Giragama, & A. Madurapperuma. 2000a. Topic-dependent Adjective Use in Japanese and Sinhala: Selection of Adjectives Differentiating Guitar Sounds. *The 7th International Work-*

shop on Human Interface Technology 27-34. Aizu: University of Aizu.

Martens, W., Y. Watanabe, & A. Marui. 2000b. User-centred Control Structure for Guitar Effects Processing: Mapping Between Perceptual Dimensions and Semantic Dimensions. *The 7th International Workshop on Human Interface Technology* 35-44. Aizu: University of Aizu.

Math Works, Inc. 2002. *Matlab Statistics Toolbox*. Natick, MA: Math Works, Inc.

Meulman, J., W. Heiser, & J. Carroll. 1986. *PREFMAP-3 User's Guide, Technical Report*. Murray Hill, NJ: Bell Telephone Laboratories.

Namba, S., S. Kuwano, T. Hashimoto, B. Berglund, Z. Rui, A. Schick, H. Hoege, & M. Florentine. 1991. Verbal Expression of Emotional Impression of Sound: A Cross-Cultural Study. *Journal of the Acoustical Society of Japan* 12.1, 19-29.

Ramsay, J. 1980. The Joint Analysis of Direct Ratings, Pair Wise Preferences, and Dissimilarities. *Psychometrika* 45, 149-165.

Rosen, V. 2003. *Computing for Non-European Languages: A Perspective from Linguistics and Southeast-Asian Languages*. Bergen: Advanced Computing in the Humanities (aco*hum), University of Bergen.

Shiffman, S., M. Reynold, & F. Young. 1981. *Introduction to Multidimensional Scaling: Theory, Methods, and Applications*. New York: Academic Press.

Wierzbicka, A. 1999. *Emotions across Language and Culture, Diversity and Universals*. Cambridge: Cambridge University Press.