Phonetica *14:* 97–107 (1966).

Some Acoustic Differences among Languages

T. D. HANLEY *, J. C. SNIDECOR * and R. L. RINGEL **

That spoken language may be described not only in terms of phonemes and morphemes, words, grammar and syntax, but also in terms of melody and prosody, is a common observation. *Halle* in 1959 ([3], p. 90) made the statement forcefully and unequivocally: "...speech evidently is a physical phenomenon and the terms in which we describe it must, in the final analysis, be translatable into the language of physics, e.g., into frequency, amplitude, phase relations of sine waves."

Heard (i.e. not measured) tonal differences among languages have been reported by *Jones*[6] in 1909 and *Kingdon* in 1958([7], pp. 266–267) among others. Observations of this type, confined to a single language, were reported as early as 1775, as noted by *Pike* ([10], p. 4).

Instrumental analyses undertaken to reveal differences between dialects and languages have been relatively few in number. *Atherton* and *Gregg*[1] in 1927 reported a pilot study of American dialect differences, and *Hanley*[4] described research in the same area in greater depth in 1951. *Parmenter* and *Blanc*[9] earlier had published results of an acoustic comparison of French and English, concluding, in part, that "In the reading of the French passage, pitch is more important as an element of accent than intensity. In the English reading, on the contrary, intensity is the more important element." Also, "French is characterized by a greater variation in pitch, 41.6% greater than in English".

Currently a large-scale project is being conducted on the University of California, Santa Barbara, campus by *Pierre Delattre* in which differences among English, Spanish and German are inves-

* University of California, Santa Barbara, California.

** University of Wisconsin; formely University of California, Los Angeles, Laryngeal Research Laboratory.

tigated by spectrographical analysis, electronic synthesis of artificial speech and motion picture X-rays.

This present investigation was designed not only to yield general acoustic data for the three language groups examined, Spanish, Japanese, and American English, but also to test a method. The attempt was made to discover whether significant differences in pitch, intensity and time factors among the language groups would be revealed by instrumental analysis. The utility of such findings was stated by *Pike*[10]: "...instrumental investigations can be very valuable to the auditory analyst, by testing his tentative descriptions to see if the vague descriptive statements of the physical basis of the pattern correspond with the facts as measured precisely."

Method

The collection of data and analysis thereof included the following:

A. Selection of subjects who were native born speakers of Spanish, Japanese, and American English;

B. Recording of subjects under acoustically controlled conditions, each reading a propositional prose passage written in his native language, and responding briefly in his own language to a question designed to elicit spontaneous speech of a like propositional nature.

C. Analysis of recordings, both reading and speaking samples, for the following:

(1) Median pitch level*.
(2) Pitch sigma**.
(3) Mean sound pressure level (SPL) above an arbitrary reference.
(4) SPL sigma*** and
(5) Phonation/time ratio, a fraction in which total reading-speaking time is the denominator and time spent phonating is the numerator; and

D. Statistical analysis for significance of the group averages calculated in C, above.

Selection of Subjects

All the subjects used in this experiment were students at the University of California, Los Angeles. The eight young men who comprised the American English-speaking group were volunteers from an introductory course in public speaking. Japanese-speaking and Spanish-speaking students were volunteers from an "English for Foreign Students" course or were recruited by means of a letter addressed to them from files provided by the Foreign Students' Advisor at the University. Table I is a summary of the demographic

* Following convention the data for the fundamental frequency parameter is presented in terms of its psychological correlate, pitch.

** Pitch sigma: For each subject, the standard deviation of his distribution of vocal frequencies measured.

*** SPL sigma: For each subject, the standard deviation of his distribution of sound pressure level peaks.

Table I

Demographic data for subject groups

Language group	Mean age*	Years residence United States	Years study English	Years residence home city
Spanish	19.5	1.0	3.5	18.0
Japanese	29.0	1.0	2.5	19.0
American English	20.5	20.5		15.0

* All mean values rounded to nearest half-year.

data collected on the three eight-person subject samples. It may be seen that the American students had spent their entire lives in the United States, but not all resided continually in the same city. However, all had been continuous residents in the same, General American, dialect region. In contrast, the Spanish-speaking students had tended to live continually in one city before coming to the United States. Five countries were represented in the Spanish-speaking sample: Mexico (3), Nicaragua (2), Ecuador, Argentina and Peru. The Japanese students were, on the average, considerably older than the other subjects and were the least fixed with respect to length of residence in their home city. Both the Spanish and Japanese groups were alike with respect to the duration of residence in the United States and the length of their exposure to the English language.

Recording of Subjects

Upon entering the recording room, each subject was given an opportunity to acquaint himself with the material prepared for him to read, the frequently used Rainbow Passage (*Fairbanks*[2], 1940) translated into his native language. The English form of the passage is as follows:

> When the sunlight strikes raindrops in the air, they act like a prism and form a rainbow. The rainbow is a division of white light into many beautiful colors. These take the shape of a long round arch, with its path high above, and its two ends apparently beyond the horizon. When a man looks for something beyond his reach, his friends say he is looking for the pot of gold at the end of the rainbow.

Upon the conclusion of this familiarization period, he was instructed to commence reading the passage aloud in a "normal oral reading manner". He was further instructed that upon the conclusion of his reading he was to commence speaking spontane-

ously in his native language about his future career plans, using a "normal conversational manner", for about one minute.

Recordings were made in a sound treated booth (I.A.C., 1200 series) whose ambient noise level was 40 dB SPL for the 0.100 – 8 KC range as measured on the "C" scale of an audio frequency spectrometer (*Bruel* and *Kjaer*, 2112). Each recording was made at a constant "mouth-microphone" distance of 18 in. through a condenser microphone-cathode follower system (*Bruel* and *Kjaer*, 4134 and 2615) and a dual channel tape recorder (*Crown*, 800). Prior to recording the speech samples, a reference tone (60 dB SPL 1 KC) was recorded free-field at the "speaker-microphone" distance of 18 in. The input controls of the recording system were adjusted as per this reference tone and all subsequent speech sample recordings were made at these settings.

Analysis of Recordings

As mentioned above, several acoustic analyses were performed on the recordings made of the subjects. Median pitch level and pitch sigma were data collected on the Fundamental Frequency Indicator (FFI) installed at the University of Florida*. A comparatively new, non-commercial instrument, FFI has been shown by *Hollien* and *Norris*[5], to have the advantage of great rapidity in data processing and most satisfactory reliability and validity.

The recorded sample sentences also were coupled from the tape recorder on which they were originally recorded to a high speed level recorder (*Bruel* and *Kjaer*, Model 2304), and average peak level measures were obtained from the resultant graphic recordings. The average, or mean, peak level was determined by a "peak counting" method which consisted essentially of measuring the amplitude of energy peaks above a pre-set reference base line (in the present investigation a 50 dB potentiometer was used and the base line was set at 60 dB SPL). Only those peaks which represented a difference of at least 5 dB from the preceding peak were counted. The two more senior authors independently measured the amplitude of peaks and found the method to be reliable.

Finally, measures of phonation/time ratio were obtained through the use of the Speech Time Analyzer (*Grason-Stadler*,

* Through the courtesy of the Communication Sciences Laboratory, University of Florida, under National Institute of Neurological Diseases and Blindness grant NB-04398, this portion of the analysis was accomplished.

E4446A). For the numerator in this fraction the analyzer was set to cumulatively measure and record the length of time that the input signals (speech) were above an arbitrarily selected reference level of 48 dB SPL. The 48 dB level was set by appropriate adjustments of the Analyzer's calibrated "threshold control system" and was facilitated by the presence of the previously noted recorded reference signal. The denominator in the fraction was the total duration of a subject's recorded sample.

Statistical Analysis

The statistical analysis of the data involved calculation, as indicated earlier, of arithmetic means and medians, where appropriate, and standard deviations. The F-statistic was used in a series of single-factor analyses to test for the significance of differences noted among the means. Calculation of the F was based on the *Winer* ([12], p. 56) Single-Factor Model 1. As revealed below, the writers also made extensive use of graphic models in evaluating the results of the experiment.

Results and Discussion

The results of statistical and graphical analyses are shown in Tables II and III and Fig. 1 and 2. As will be observed at the outset, the tables are more compressed than is typical in cases where statistical significance is tested. The compression was done to conserve space and focus attention on the mean values obtained. The full statistical treatment is available on request.

As revealed in Table II and the upper row of cubes in Fig. 1, distinct differences were found in the acoustic data from the three language groups. A rather liberal level of confidence (10%) having been adopted for this small sample pilot study, all acoustic para-

Table II

Statistical analysis of data from reading sample

	Group results			Significance	
	Spanish	Japanese	American English	F	P
Median pitch level (cps)	124.6	130.6	105.6	3.81	0.05
Pitch sigma (semitones)	2.8	3.5	3.4	2.90	0.10
Mean SPL (dB above ref.)	7.3	8.5	10.6	13.42	0.01
SPL sigma (dB)	2.9	2.8	2.3	3.85	0.05
Phonation/time ratio	0.60	0.56	0.63	3.20	0.10

meters were found to be significant in the data collected from the reading sample. That is, differences among language groups for pitch level and sigma, sound pressure level and sigma, and phonation/time ratio were of sufficient size that their occurrence by chance was unlikely.

The method chosen for representing these data graphically, construction of cubes for each language group with relative height, width, and depth representing pitch, sound pressure level, and phonation/time ratio, respectively, makes possible rapid and interesting comparisons. For example, the upper row of projections, for the reading passage, shows the Spanish and Japanese groups to be higher in pitch level and lower in sound pressure level than the American English group, with phonation/time ratio distributed at rather even intervals from the Japanese low of 0.56 to the American English high of 0.63.

Somewhat the same observations can be made for the spontaneous speaking sample, Table III and the lower row of cubes in Fig. 1, except that positions for Japanese and Spanish groups on the phonation/time ratio analysis are reversed. Moreover, a qualification must be noted. On three of the analyses, for pitch level, sound

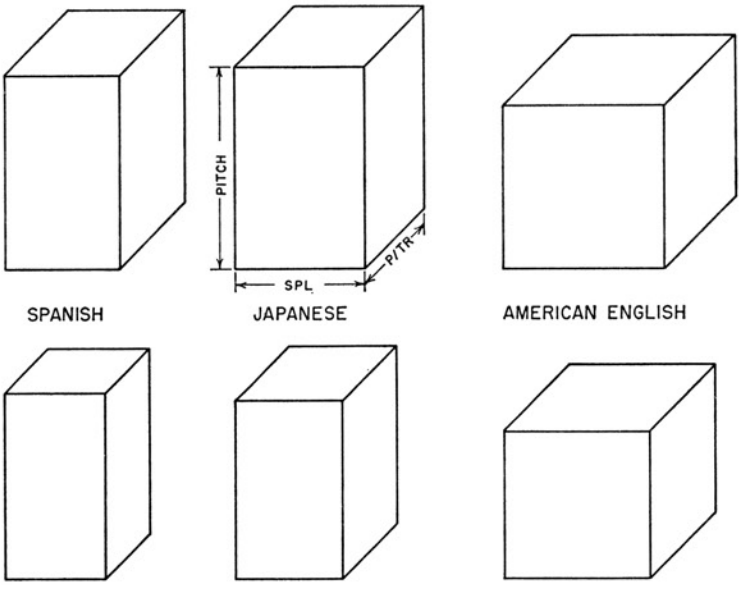

Fig. 1. Comparison of oral reading performances (above) and speaking performances (below) on three vocal parameters by three language groups.

Table III

Statistical analysis of data from speaking sample

	Group results			Significance	
	Spanish	Japanese	American English	F	P
Median pitch level (cps)	121.4	116.8	95.3	4.60	0.05*
Pitch sigma (semitones)	2.6	3.2	2.4	3.16	0.10
Mean SPL (dB above ref.)	6.4	7.1	9.3	8.29	0.01
SPL sigma (dB)	2.7	2.7	2.0	3.92	0.05*
Phonation/time ratio	0.39	0.48	0.61	6.20	0.01*

* On the basis of the F_{max} statistic (Winer[12]), the hypothesis of homogeneity of variance would be rejected.

pressure level, and phonation/time ratio, the hypothesis of homegeneity of variance had to be rejected, thus casting the validity of the results into serious question. The writers believe that acceptance and analysis of spontaneous speaking samples as short as 15 sec accounts for the variability within groups that showed up in the F_{max} test for homegeneity of variance. In spite of this statistical weakness, the speaking data are shown to be remarkably similar to the reading data, except for a general reduction along axes of the cubes, a finding consistent with researches performed earlier.

The variability data from the experiment are shown in more conventional graphical form in Fig. 2 and summarized in Table III. There the Japanese and American groups are seen to be quite similar in pitch variability (left bars) and greater than the Spanish. In contrast, the Spanish and Japanese groups were more alike in variability around mean sound pressure level (right bars), with the

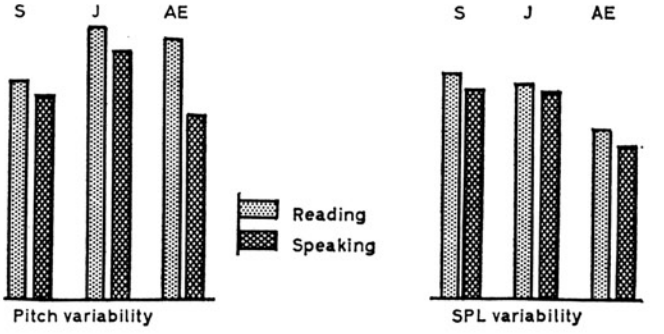

Fig. 2. Vocal variety exhibited in oral reading and speaking performances by three language groups.

American group somewhat lower in this statistic. Variability in both parameters was lower in the speaking than the reading sample, conspicuously so for pitch sigma in the American English group. This finding is consistent both with previous researches and with the common statistical correlation between magnitude of the mean and the standard deviation.

It is to be noted that no tests of significance were applied to the apparent differences between reading and speaking samples, such results being secondary to the primary objectives of the research.

However, certain speculations are in order as regards differences between the speaking and reading performances.

First, the reading performance appears inherently "smoother" and more consistent than the speaking performance which no doubt accounts for the greater reliability noted among the measures in Table II (Reading) as constrasted with the measures in Table III (Speaking). The reliability of the statistics for speaking would no doubt be stabilized with larger samples.

Second, the relationship of phonated time to total time gives some measure of the relative legato *vs* staccato characteristics of speech in a reading performance in which each performer has an opportunity to become familiar with the materials to be read to the end of developing a relative effective performance. Each speaker in this study used relatively more phonated time in reading than in speaking, a result consistent with previous research.

Third, perusal of Tables II and III indicates that the reading performance was consistently higher in pitch than the speaking performance: in the case of Spanish by a difference of only 3.2 cps, but in the case of Japanese, 13.8 cps, and in American English, 10.3 cps. This apparent trend is consistent with a finding noted by *Snidecor*[11] who, using identical materials for reading and speaking with General American male performers, noted that mean pitch levels for reading and speaking were 132 cps and 120 cps, respectively, a difference of 12 cps, or 0.89 of a tone. Again the differences in this study are consistent with previous findings.

Fourth, as previously noted, the higher pitched reading performance in each language group was more variable than the lower pitched speaking performance as indicated by higher sigmas for the former. This is consistent with findings by *McIntosh*[8] who indicates that higher pitched performances are more flexible in pitch usage, whereas more flexible performances are higher pitched. However,

pitch level and flexibility seem not necessarily to be related when comparisons already mentioned among the three languages are made. For example, American English was lower pitched than Spanish, but more variable in reading, less so in speaking. It is likely that variability differences are consistent only between the speaking and reading performances of individual speakers.

As regards the methodology of this study, it is to be noted that accurate machine analysis of pitch takes only a fraction of the time needed for phonellographic analysis and gives, for most purposes, comparable results. The computation of phonation/total time, once laborious, has become almost automatic. Measurement of intensity values remains time consuming, but in the future is surely amenable to some type of computer processing. The methods here utilized are, of course, recommended for further testing and validation as research tools for the linguist as well as for the experimental phonetician.

Conclusions

Reading and spontaneous speaking samples were collected from three groups of subjects, each composed of eight young men. The groups were native speakers of Spanish, Japanese, and American English. Acoustical and statistical analyses were performed on these samples with the following results:

A. In the reading sample, statistically significant differences among groups were found for pitch level and variability, sound pressure level and variability, and phonation/time ratio;

B. In the spontaneous speaking sample the same results were found, although heterogeneity of variance in pitch level, sound pressure level variability, and phonation/time ratio variability sharply limits the conclusions that can be drawn from these data;

C. Graphical representations in cube form of the data collected in this study appear to provide an easily assimilated set of comparison, viz.,

(1) Male speakers of Spanish and Japanese are alike in pitch level and higher than speakers of American English;

(2) the same two groups are alike in sound pressure level above an arbitrary reference and lower than speakers of American English;

(3) male speakers of American English spend proportionately greater time in phonating than speakers of Japanese and Spanish;

D. Graphical comparisons in standard bar graph form reveal the Japanese and American English speakers to read with greater pitch variability than speakers of Spanish, whereas the speakers of Spanish and Japanese use greater sound pressure level variability than speakers of American English.

E. Although not in the main stream of the study, it was observed that there was relatively greater phonated time in reading than in spontaneous utterance. The reading performances were higher in pitch and more variable in pitch usage than the speaking performances. These findings are consonant with previous comparable research.

Finally, with respect to this general problem as posed in the introduction, the instrumental-analytical-graphical method employed in this research does appear to have merit as a means of describing and differentiating among spoken languages.

Summary

Modern instrumental analysis, within the limits of the experiment, and for the more stable oral reading performance, indicates differences among the three language groups as follows: On a pitch continuum from high to low, mean pitch level orders Japanese, Spanish, and American English in that order. On a pitch variability continuum, Japanese is most variable, with American English and Spanish following. American English proved to have more intensity with Japanese and Spanish following in that order. In intensity variability, Spanish was most variable with Japanese and American less variable. The highest phonation/time ratio favored American English with Spanish and Japanese following. Although not in the main stream of the study, there was relatively greater phonated time in reading than in speaking. The reading performances were higher in pitch and more variable in pitch usage than were the speaking performances.

Einige akustische Unterschiede der Sprachen

Zusammenfassung

Eine moderne instrumentelle Analyse gesprochener Sprachen liefert – in den experimentell bedingten Grenzen und bei Berücksichtigung nicht zu stark variierender Lesungen – folgende Unterschiede zwischen den unteruchten Sprachen:

Auf einer Skala der mittleren Tonhöhe besteht folgende Ordnung – von hoch nach tief: Japanisch, Spanisch, amerikanisches Englisch. Die Variabilität der Tonhöhe ist beim Japanischen am größten; es folgt das Amerikanische, dann das Spanische. Das

Amerikanische erweist sich als stärker hinsichtlich der Intensität als das Japanische und Spanische – in dieser Reihenfolge. Die Variabilität der Intensität war im Spanischen am größten, es folgen das Japanische und das Amerikanische. Das Verhältnis Phonation : Zeit war im Amerikanischen am größten, es folgte das Spanische, dann das Japanische. Beiläufig zeigte sich, daß beim Lesen mehr Zeit zur Phonation verwendet wird als beim freien Sprechen. Das Gelesene wurde mit höherer Stimme gesprochen, und die Tonhöhe schwankte auch stärker als bei frei Gesprochenem.

Quelques différences acoustiques des langues

Résumé

L'analyse instrumentale effectuée par des procédés modernes dans les limites nécessaires d'une expérimentation et avec des variations minimes entre les différentes lectures fait ressortir les différences suivantes dans les trois langues étudiées:

Dans l'ordre décroissant de la hauteur musicale on a successivement le japonais, l'espagnol et l'anglo-américain. C'est en japonais que les variations de hauteur sont les plus accusées; viennent ensuite l'américain, puis l'espagnol. L'américain se révèle comme ayant une intensité plus forte que le japonais qui est suivi de l'espagnol. Les variations de l'intensité sont les plus marquées en espagnol, elles sont moindres en japonais et en américain. Le rapport phonation: temps nous est apparu comme étant le plus élevé en anglo-américain qui est suivi dans l'ordre par l'espagnol et le japonais. On a constaté par ailleurs que le temps de phonation est plus long dans la lecture que dans le discours libre. Un texte lu a été dit sur un registre plus élevé et les changements de hauteur y ont été plus prononcés que dans le discours libre.

References

1. *Atherton, H.* and *Gregg, D.:* A study of dialect differences. American Speech. *4:* 216–223 (1927).
2. *Fairbanks, G.:* Voice and articulation drillbook (Harper & Brothers, New York and London 1940).
3. *Halle, M.:* The sound pattern of Russian (Mouton and Co., 's-Gravenhage 1959).
4. *Hanley, T.:* An analysis of vocal frequency and duration characteristics of selected samples of speech from three American dialect regions. Speech Monographs. *18:* 78–93 (1951).
5. *Hollien, H.* and *Norris, R.:* Pitch characteristics of adult males: an assessment of FFI (Progress report submitted to the National Institute of Neurological Diseases and Blindness. July 1, 1963).
6. *Jones, D.:* Intonation curves (Leipzig and Berlin 1909).
7. *Kingdon, R.:* The groundwork of English intonation (Longmans Green and Co., London 1958).
8. *McIntosh, C.:* A study of the relationship between pitch level and pitch variability in the voices of superior speakers (Unpublished Ph. D. Dissertation, State University of Iowa, 1939).
9. *Parmenter, C.* and *Blanc, A. V.:* An experimental study of accent in French and English. PMLA. *48:* 598–607 (1933).
10. *Pike, K.:* The intonation of American English (University of Michigan Press, Ann Arbor 1945).
11. *Snidecor, J.:* A comparative study of the pitch and duration characteristics of impromptu speaking and oral reading. Speech Monographs. *10:* 50–56 (1943).
12. *Winer, B.:* Statistical principles in experimental design (McGraw-Hill, New York 1962).

Authors' addresses: Drs. T. D. Hanley, J. C. Snidecor, University of California, *Santa Barbara, Calif.* (USA); Dr. R. L. Ringel, University of Wisconsin, Speech Sciences Laboratory, *Madison, Wisconsin* (USA).