Acoustical Analysis of Spanish Vowels Produced by Laryngectomized Subjects

Teresa Cervera
José L. Miralles
Universitat de València
València, Spain

Julio González-Àlvarez
Universitat Jaume I
Castellón, Spain

The purpose of this study was to describe the acoustic characteristics of Spanish vowels in subjects who had undergone a total laryngectomy and to compare the results with those obtained in a control group of subjects who spoke normally. Our results are discussed in relation to those obtained in previous studies with English-speaking laryngectomized patients. The comparison between English and Spanish, which differ widely in the size of their vowel inventories, will help us to determine specific or universal vowel production characteristics in these patients. Our second objective was to relate the acoustic properties of these vowels to the perceptual data obtained in our previous work (J. L. Miralles & T. Cervera, 1995). In that study, results indicated that vowels produced by alaryngeal speakers were well perceived in word context.

Vowels were produced in CVCV word context by two groups of patients who had undergone laryngectomy: tracheoesophageal speakers (TES) and esophageal speakers. In addition a control group of normal talkers was included. Audio recordings of 24 Spanish words produced by each speaker were analyzed using CSL (Kay Elemetrics). Results showed that F1, F2, and vowel duration of alaryngeal speakers differ significantly from normal values. In general, laryngectomized patients produce vowels with higher formant frequencies and longer durations than the group of laryngeal subjects. Thus, the data indicate modifications either in the frequency or temporal domain, following the same tendency found in previous studies with English-speaking laryngectomized speakers.

KEY WORDS: speech perception, vowel recognition, laryngectomized voice

Most of the acoustic analysis of alaryngeal speech has been carried out with an English-speaking sample during the production of vowels. The early literature on voice production in laryngectomized subjects focused on those characteristics related to the voicing source, and it was assumed that the rest of the vocal tract behaved substantially the same as in normal voice users (Damste, 1958). However, later studies in which acoustical analysis of vowels was conducted suggested that alaryngeal speakers might also have altered vocal cavity transmission characteristics. In these studies, vowels consistently showed higher formant frequencies for alaryngeal talkers, both tracheoesophageal (TES) and esophageal, than for normal speakers. Sisty and Weinberg (1972) studied English vowels in [hVd] sequences produced by male and female esophageal talkers and found consistently higher formant values. Their results were similar to those previously obtained for Finnish vowels produced by male esophageal speakers in a CVC context (Kytta, 1964). TES talkers have shown higher vowel

1092-4388/01/4405-0988

formant frequencies than normal and esophageal speakers (Sisty & Weinberg, 1972). The results of these studies lead to the hypothesis that total laryngectomy results in a shortened vocal tract relative to normal subjects (Christensen & Weinberg, 1976; Sisty & Weinberg, 1972). This hypothesis seems to be compatible with some physiological data presented by Diedrich and Youngstrom (1966). These authors obtained cinefluorograms of a patient before laryngectomy during the production of some vowels and compared the results with those obtained after laryngectomy. It was found that the effective vocal tract length was reduced after the surgical operation. Unfortunately, few physiological studies with laryngectomized subjects have been conducted to provide evidence about the relations between the vocal tract configuration and the resulting acoustic characteristics. This likely is due to the relative inaccessibility of the relevant vocal tract areas during speech production.

The literature also has shown temporal differences between normal and alaryngeal speakers of the English language. Robbins, Fisher, Bloom, and Singer (1984) found that laryngectomized patients using TES or esophageal voice produced speech at slower rates, measured as words per minute, and with greater total duration for a paragraph reading than normal subjects. Christensen and Weinberg (1976) and Diedrich and Youngstrom (1966) also reported longer vowel duration in excellent esophageal talkers. The study by Robbins, Christensen, and Kempster (1986) showed longer vowels for TES and esophageal talkers than for the normal subjects. At the same time no significant differences were found between the TES and esophageal groups. The hypothesis that these authors proposed to explain these temporal disturbances is that alaryngeal patients have poorer motor control of the neoglottis than a laryngeal speaker's motor control of the glottis during phonation. The limited control of the resistance offered by the cricopharyngeal segment may be responsible for the difficulty in initiating and terminating voicing in these patients; thus, duration of the phonetic segment may be altered (Moon & Weinberg, 1987).

Thus, both frequency and temporal characteristics of vowels produced by laryngectomized speakers seem to be altered, compared to those of normal speakers. In spite of these modifications, vowel intelligibility rates in word context are around 90% according to our previous perceptual study (Miralles & Cervera, 1995). In that study, the same list of Spanish words produced by the TES, esophageal, and normal speakers in the present study were presented to a group of listeners in a perception task. Phoneme confusion scores were used to construct confusion matrices. Significant differences were found among phoneme classes in both groups of alaryngeal speakers. Spanish vowels produced by laryngectomized speakers were perceived significantly better than the other phoneme classes (i.e., fricatives, stops, nasals, and glides). We attribute this finding to the simplicity of the Spanish vowel system and inferences made by the listeners about vowel identity in word contexts.

The purpose of the present study was (a) to obtain formants (F1 and F2) and duration values in stressed vowels in CVCV context in patients who had undergone laryngectomy and who used TES and esophageal voice, and (b) to compare the results to those obtained in a group of normal speakers. Although other acoustical parameters seem to play a role in the perception of vowels, the first two formants (Delattre, Liberman, Cooper, & Gerstman, 1952; Jones, 1960; Klein, Plomp, & Pols, 1970; Peterson & Barney, 1952) and duration (Ainsworth, 1972; Lehiste & Peterson, 1961; Peterson & Lehiste, 1960) have been shown to be the most relevant parameters in vowel recognition.

In the present study vowels in word (CVCV) contexts were used with the objective of reproducing speech as natural sounding as possible. Stressed vowels were selected because one of our objectives was to compare the results of our study with those conducted with alaryngeal English speakers. In the latter studies CVC or [hVd] contexts were used. In Spanish, CVCV is a more common structure than CVC. Thus, although syllabic structure was not the same, in both cases the stressed vowel was analyzed. Words were used because it is believed that the deviant aspects of pathological voices will be better manifested in words or longer passages than in isolated vowels, as coarticulation involves higher physiological complexity than isolated vowels.

Method

Stimuli and Subjects

Twenty laryngectomized subjects participated in the experiment. Ten of the patients had undergone tracheoesophageal shunt (TES) and used tracheoesophageal voice as their primary mode of communication, and the 10 subjects without TES used esophageal voice. All the patients came from the Service of Otolaryngology of the Hospital Clinico Universitario in València, Spain. TES patients did not receive any specific speech rehabilitation. Esophageal patients received instructions by volunteers from the Association of Laryngectomized Patients in València about the air injection method. All the subjects spoke standard Castillian Spanish. All subjects were male, and their ages ranged from 51 to 64 years, with a mean age of 59 years. In the group of TES patients, post-operation time ranged from 6 to 13 months, with a mean of 10 months. In the group of esophageal

patients without the shunt, the post-laryngectomy time ranged from 17 to 52 months (M = 32.5). The two groups of patients could not be matched for the period of post-operation because during the last years all the laryngectomies were accompanied by TES in our hospital. Before this time period, none of the patients had undergone any TES and all the laryngectomized patients had to use esophageal voice. TES patients used the Herrmann prosthesis (Herrmann, 1978). Choice of this prosthesis was made by the surgeons. This one-way valve has shown aperture pressures considered within the range of medium to low values (Gimenez et al., 1993). All of the TES speakers used finger occlusion of the tracheostoma to produce voice. All of the alaryngeal speakers, TES and esophageal, were judged to have acceptable speech by a group of speech therapists. None of the patients exhibited any other speech pathology.

A control group of 10 normal-voiced subjects produced and recorded the same list of words. These subjects were male and were of similar ages to subjects in the patient groups.

Procedure

Twenty-four Spanish two-syllable (CVCV) words produced by 20 laryngectomized male patients were recorded. The list of words used in this study and our previous perceptual study (Miralles & Cervera, 1995) is one of the bisyllabic lists used in audiometric tests of the Spanish population (Perelló & Mas-Dalmau, 1980). The phonetic characteristics of these words reflect the most frequent syllabic structure and stress patterns in the Spanish language (see Appendix). The vowels analyzed were those in the stressed position (first syllable).

A Revox B77 tape recorder and a Sennheizer microphone were used in this experiment. Each subject's speech was recorded individually in an audiometric booth. A Sennheizer microphone was mounted at a distance of 15 cm from the mouth. The subjects were instructed to read the list of words at their normal intensity with an interval of approximately three seconds between consecutive words. Those words produced by alaryngeal speakers that couldn't be recognized in the perceptual study in our previous work (Miralles & Cervera, 1995) were eliminated from this study.

Acoustic Analysis

The recorded lists of words were digitized at a 20 kHz sampling rate using a CSL (Kay Elemetrics Corp., model 4300B) 14 bit resolution card and were stored on an 486 IBM computer hard disk for subsequent analysis. An anti-aliasing filter with a 10 kHz cutoff frequency was used before the A/D conversion. A preemphasis with a factor of 0.8 was applied.

All the analysis was done using the CSL software. Broadband spectrograms (300 Hz) of each word of the list produced by each speaker were obtained and used to facilitate the identification of the vowel. Once this segment was identified, LPC and FFT spectra were computed for the vowel segments using 256 consecutive samples centered in the middle of the vowel segment. The signal was then multiplied by a 45 ms Hamming window, and LPC coefficients were computed using an autocorrelation method (Atal & Hanauer, 1971). The LPC order was 23 according to the criterion of Rabiner and Shaffer (1978). Vowel duration was measured as the steady-state portion in the first syllable of the CVCV sequence. The starting and ending points of the vowels were measured by hand from visual inspection of high-resolution spectrograms. The time period between these two points showed no formant movement.

This procedure differs from that used in the Robbins et al. (1986) study and from the standard procedure suggested by Peterson and Lehiste (1960), which has been used in various studies measuring vowel duration in normal speakers (Hillenbrand, Getty, Clark, & Nearey, 1995; Hillenbrand & Nearey, 1999). In those studies vowels were presented in [hVd] context (Hillenbrand et al., 1995; Hillenbrand & Nearey, 1999; Peterson & Lehiste, 1960) or in /pik/, /kap/, and /kup/ sequences (Robbins et al., 1986). The starting and ending points of the vowels were determined by visual inspection of vertical striations that correspond to phonation time in the spectrograms. This method of vowel measurement includes transitions. In the present study, however, vowels were presented in all consonant contexts, including liquids and nasals, which are also characterized by formants and their transitions. The criterion of Peterson and Lehiste (1960) could not be applied, so the steady-state time was used as a criterion. Total word duration was also measured.

Formant frequencies were measured at the midpoint of the vowel. Detection of a formant was based on the visual examination of LPC and FFT spectra. The first two peaks from the LPC spectrum were selected as F1 and F2 values. General knowledge of acoustic phonetics also helped in the process of estimating vowel formants (Ladefoged, 1967). First and second formant frequencies were measurable 97% of the time in pathological groups and 100% of the time in the normal group. Third formant frequencies were measurable less than 50% of the time in both groups of patients and were not included in our results. Similar difficulties in measuring third formants in pathological voices are reported in the literature (Sisty & Weinberg, 1972). In our study, TES and esophageal voices presented similar difficulties in measuring higher formants; however, presence of noise was more evident in TES than in the esophageal group.

Reliability

All the measures were obtained separately by two different investigators, and reliability measures were obtained afterwards. Reliability of vowel duration measurements was obtained by computing the average absolute difference expressed in percentages between the measures made by one investigator and those made by a second investigator. This reliability estimate has been used in other studies involving duration measurements in normal speech (Allen, 1978; Hillenbrand et al., 1995; Smith, Hillenbrand, & Ingrisano, 1986) and in speech-language pathology research (Onslow, van Doorn, & Newman, 1992; Swanson, Leonard, & Gandour, 1992). The absolute difference between investigators measure was 7.8%.

To estimate the reliability of the formant frequency extraction, we used the average absolute differences, expressed in percentages, between the measures in formants frequency obtained by the two investigators (Di Benedetto, 1989). For F1 the result was 1.9%, and for F2 the result was 1.5%.

Results

Vowel Duration

Means of vowel duration for the five Spanish vowels produced by TES, esophageal, and normal speakers are presented in Table 1. We can observe that TES patients had the longest vowel duration, whereas esophageal patients had values that were more similar to those of normal speakers. A two-way ANOVA, with group (TES, esophageal, and normal) and vowel (/i/, /e/, /a/, /o/, and /u/) as independent variables and vowel duration as a dependent measure, showed main effects for group (MS = 154844.63, F = 106.1, $p < 0.001$) and vowel (MS = 2477, F = 1.69, $p < 0.001$). The effects for group ´ vowel interaction were not significant. Post hoc Tukey test showed significant differences between TES and esophageal ($p < 0.0001$), TES and normal ($p < 0.0001$), and esophageal and normal ($p < 0.0001$).

Formant Frequencies

Mean values for the first and second formants of the vowels produced by TES, esophageal, and normal voices are presented in Table 1. Comparisons between alaryngeal and normal speakers showed that, with the exception of mean F2 value for /o/ in the esophageal group, F1 and F2 had higher values in TES and esophageal groups than in the normal group. The differences in F1 values in the pathological and normal groups were around 60 Hz. Greater differences were found in F2 (160 Hz, on the average), especially for the vowel /i/.

MANOVA analysis was performed with F1 and F2 as dependent measures and group (TES, esophageal, and normal) and vowel (/i/, /e/, /a/, /o/, and /u/) as factors. The main effects for group were significant for both (F1: MS = 7398, F = 14.42, $p < 0.001$; F2: MS = 66397.2, F = 11.16, $p < 0.001$). The main effects for vowel were also significant for both (F1: MS = 7398.1, F = 293, $p < 0.001$; F2: MS = 66397.2, F = 504.5, $p < 0.001$). The interaction

Table 1. Mean vowel durations (in ms) and F1 and F2 frequency (Hz) of Spanish vowels produced by 10 tracheoesophageal (TES), 10 esophageal, and 10 normal subjects.

		TES		Esophageal		Normal	
Vowel	Parameter	*M*	*SD*	*M*	*SD*	*M*	*SD*
/i/	F1	355	81	348	155	331	41
	F2	2636	554	2518	439	2241	119
	duration	139	40	11	25	81	26
/e/	F1	546	69	532	66	502	75
	F2	2087	296	2180	271	1872	111
	duration	142	61	99	34	78	28
/a/	F1	781	94	866	100	718	86
	F2	1576	207	1498	257	1479	75
	duration	138	53	97	33	89	23
/o/	F1	603	64	540	120	533	73
	F2	1079	78	970	147	1156	418
	duration	131	51	89	36	83	19
/u/	F1	459	44	326	100	376	30
	F2	883	54	832	136	773	116
	duration	128	55	95	31	75	40

effects for group × vowel were significant for F1 (MS = 7398, F = 5.9, $p < 0.001$) and for F2 (MS = 66397.2, F = 6.4, $p < 0.001$).

Tukey post hoc analysis conducted to test differences among levels of the group factor in both dependent measures, F1 and F2, showed significant differences for TES and esophageal ($p < 0.003$), TES and normal ($p < 0.0001$), and esophageal and normal ($p < 0.0001$) for F1. For F2, Tukey post hoc test showed significant differences for TES and esophageal ($p < 0.006$) and TES and normal ($p < 0.022$), but no significant differences between esophageal and normal speakers.

The relation among vowels in the three groups of talkers is shown in Figure 1. Plots of F1 and F2 mean values for the five vowels in the three groups of talkers are presented. It can be seen that in spite of the absolute frequency formant differences between the three groups of talkers, the relative positions are maintained.

Dispersion of vowel categories with respect to F1 and F2 values are presented in Figures 2, 3, and 4 for each group. Minimal overlap is evident among vowels in the TES and esophageal groups, as in the normal group.

Word Duration

Mean word duration was 568.4 ms (SD = 168) in TES group and 399 ms (SD = 116) in the esophageal group. In the control group of laryngeal speakers the

Figure 1. F1 and F2 mean values for vowels in TES, esophageal, and normal groups.

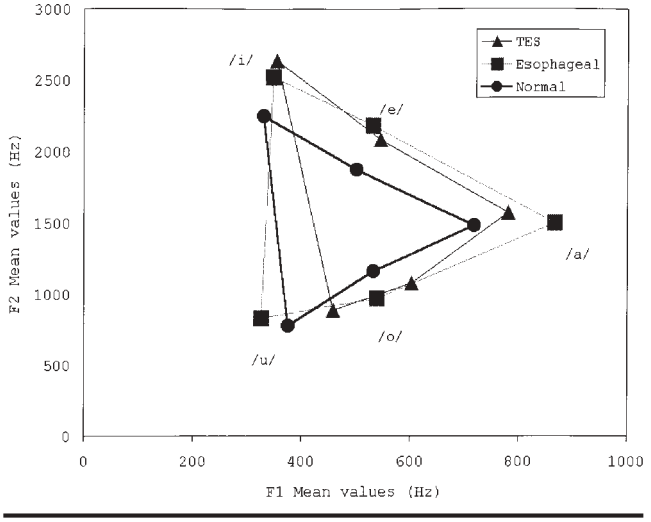


Figure 2. F1 and F2 values for vowels in the TES group.

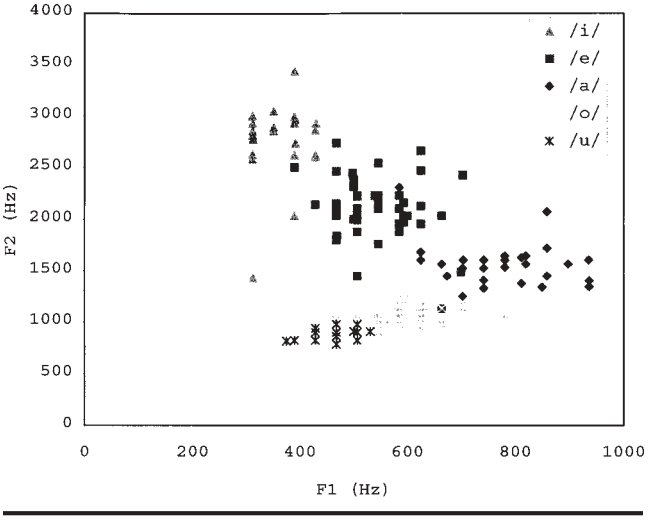


Figure 3. F1 and F2 values for vowels in the esophageal group.

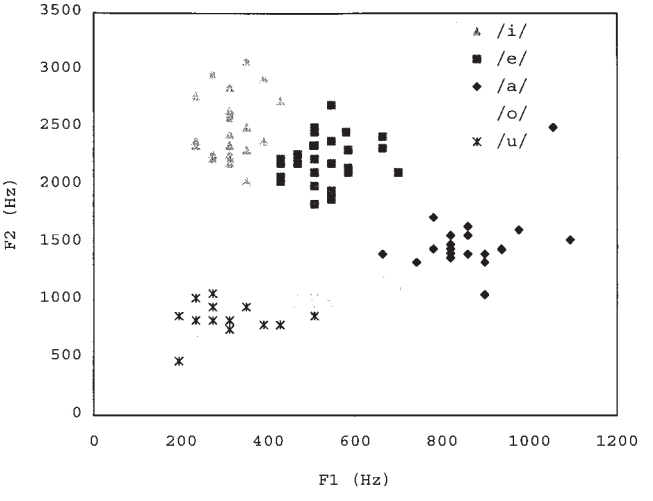
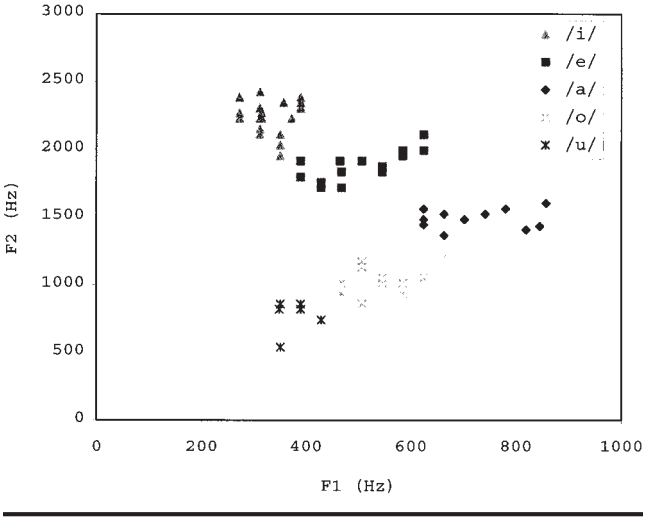


Figure 4. F1 and F2 values for vowels in the normal group.



mean duration was 434 ms (SD = 63). One-way ANOVA of word duration as a dependent measure and group (TES, esophageal, and normal) as a factor showed that the main effects were significant (MS = 1972148.5, F = 136.8, $p < 0.001$). Tukey post hoc analysis showed significant differences between TES and esophageal ($p < 0.0001$), TES and normal ($p < 0.0001$), and esophageal and normal ($p < 0.0001$) speech.

Discussion

In general, time domain analysis reveals that the two groups of laryngectomized speakers have significantly longer vowels than the control group of laryngeal speakers. The total word duration was significantly longer only in the TES group**,** whereas the esophageal group showed significantly shorter word productions than the normal group. We attribute this finding to the reduction of the second syllable, as we could determine from visual inspection of the spectrograms. The limited reservoir of air in the esophagus used to phonate may produce reduced second syllable duration compared to the first syllable of the word.

Our findings agree with most of the English language studies with laryngectomized patients. Longer durations could be attributed to either the syllable duration (Christensen & Weinberg, 1976; Diedrich & Youngstrom, 1966) or vowel duration (Robbins et al., 1986).

As in English, duration is not a phonetic feature in Spanish, and it is assumed that both speakers and listeners are not sensitive to this dimension in separating vowel categories. However, vowel duration may play a role in the process of normalization of the vocal tract of the speaker. Rosner and Pickering (1994) suggest that duration becomes an aspect of the prototypical form of a vowel speaker-specific factor, together with other factors such as F0 and higher formants.

Comparing TES and esophageal voices of Spanish speakers shows that the temporal characteristics of vowels in these patients are similar to durations found in English language studies (Christensen & Weinberg, 1976; Robbins et al., 1986). That is, the longest vowel durations were particular to the TES patients, the shortest to the normal group, and intermediate values to the esophageal group. The differences in absolute vowel duration values in previous studies with English-speaking alaryngeal subjects and the present work with Spanish alaryngeal speakers could be attributed not only to linguistic differences but also to the different contexts used. The durations for /a/, /i/, and /u/ in CV context in Christensen and Weinberg (1976) and Robbins et al. (1986) are not directly comparable to the Spanish vowels in CVCV context used in this study. In a word context the duration of the vowel is normally shorter than in the syllable environment because of the articulatory rate.

Some hypotheses suggested by Christensen and Weinberg (1976) and Robbins et al. (1986) attribute the longer vowel duration found in the productions of alaryngeal patients to the slower decay in pharyngoesophageal segment vibrations as compared with the laryngeal vibrations in the normal speakers. In alaryngeal voice the motor control of the pharyngoesophageal segment is not comparable with the motor control capabilities of the normal speaker (Moon & Weinberg, 1987). However, we agree with Sisty and Weinberg (1972) in that some caution should be used in making this hypothesis until the myoelastic properties of the esophageal sphincter are better understood.

In the frequency domain, F1 and F2 formant values obtained in our control group of normal speakers agreed, in general, with those obtained in other studies in the normal population (Bradlow, 1995; Golderos, 1984; Quilis, 1981). Differences observed among those previous studies and the present one are attributed to differences in the material employed. Golderos (1984) and Quilis (1981) used vowels produced in isolation. In Bradlow's study (1995) CVCV sequences were employed, but the target vowels occurred only between /p/, /b/, or /t/. In our study, the vowels were coarticulated with all the phoneme classes: fricatives, stops, nasals, and liquids (see Appendix). Formant structure can vary slightly due to consonantal context (Stevens & House, 1972); thus minimal differences found among different studies can be attributed to coarticulation effects.

In the present study it was evident that there were higher formant frequency values for all the vowels in both alaryngeal groups than in the normal group. In this regard, our results are consistent with previous studies carried out with English-speaking alaryngeal subjects. Sisty and Weinberg (1972) found higher formant frequency values in esophageal speakers with vowels produced in [hVd] context. In that work, the average increases for F1 and F2 in the esophageal speakers as compared with the normal speakers were 122 Hz and 325 Hz, respectively. In the present study, less difference between esophageal and normal groups was found.

F1 versus F2 diagrams show no overlap among vowel categories in the normal group, as expected. In both groups of laryngectomized subjects it was found that there was minimal or no overlap among vowel categories. This finding differentiates the present study from studies with English-speaking patients. In Sisty and Weinberg's (1972) study considerable overlap was found between /ɪ/ and /i/, with less overlap between other categories. In the normal population the differences between English and Spanish vowel systems have been

noted (Fox, Flege, & Munro, 1995). Hillenbrand et al. (1995) and Peterson and Barney (1952) showed a certain degree of overlap among adjacent vowels categories. Similar studies in Spanish language (Flege, 1991) have not shown overlap among vowel categories. This is understandable if we take into account the simplicity of the Spanish vowel system.

Mean values of the acoustic vowel space (Figure 1) showed a general upward shift, especially in the F2 dimension, in both groups of laryngectomized subjects as compared with the normal group. To the extent that articulatory interpretations of formant data are valid (Stevens & House, 1955) a general observation can be made. If we relate these two acoustic parameters with high/low and front/back phonetic dimension (Delattre, Liberman, & Gerstman, 1952; Peterson & Barney, 1952), it seems that TES and esophageal speakers articulate vowels with fronted and higher tongue positions relative to the tongue position in normal speakers. To explain the differences between alaryngeal and normal formant frequency values, the effect of shortening of the vocal tract as a result of total laryngectomy as also been mentioned (Christensen & Weinberg, 1976; Sisty & Weinberg, 1972).

This hypothesis agrees with previous findings on the perception of vowels produced by normal speakers, in which the formant frequency values of vowels produced by men, women, and children are compared. Vocal tract length seems to be the most important factor determining the positions of the formant frequencies. With shorter vocal tracts higher formants are found (Peterson & Barney, 1952).

The size of the vowel space also has been related to speaking rate. Slower speaking rates have produced expansion of the vowel space in dysarthric populations (Turner, Tjaden, & Weismer, 1995). In this regard, we can hypothesize that longer vowel productions in laryngectomized patients might also contribute to an expansion of the vowel space. Mean values for formant frequencies show this tendency clearly.

However, as can be seen in the dispersion diagrams, in TES and esophageal groups some small degree of overlap between adjacent vowel categories is observed, unlike the normal group diagram. We can attribute this finding to the fact that alaryngeal groups are less homogeneous than groups of normal speakers. Standard deviations from the mean F1 and F2 formants values are larger in alaryngeal groups than in the normal group. In spite of this overlap found among alaryngeal speakers, vowel perception was quite good according to our previous perceptual study, in which we obtained error scores below 10%. So, we can hypothesize, according to the dispersion theory (Lindblom, 1986), that vowels would tend to be maximally, or sufficiently, dispersed in the acoustic space so as to minimize perceptual confusions between vowel categories. However, future studies are needed to better understand the relation between vowel duration and expansion of the vowel space in alaryngeal speakers.

Finally, in spite of the differences in absolute formant frequency values in the two groups of alaryngeal speakers, as compared with normal values, these changes are systematic. This finding indicates that the relative positions from vowel to vowel are maintained, and, thus, perception can be accomplished. In this regard, many researchers agree that, rather than absolute values of F1 and F2, listeners use various combinations of the acoustic information, such as F1-F0 Bark distance (Fahey & Lopez-Bascuas, 1994) and F2-F1 and F3-F2 (Fox et al., 1995). Thus, the auditory system makes adjustments for the differences in vocal tract length, fundamental frequency, and so forth across speakers.

References

Ainsworth, W. A. (1972). Duration as a cue in the recognition of synthetic vowels. *Journal of the Acoustical Society of America, 51,* 648–651.

Allen, G. D. (1978). Vowel duration measurement: A reliability study. *Journal of the Acoustical Society of America, 63,* 1176–1185.

Atal, B., & Hanauer, S. L. (1971). Speech analysis and synthesis by linear prediction of the speech wave. *Journal of the Acoustical Society of America, 50,* 637–655.

Bradlow, A. R. (1995). A comparative acoustic study of English and Spanish vowels. *Journal of the Acoustical Society of America, 97*, 1916–1924.

Christensen, J. M., & Weinberg, B. (1976). Vowel duration characteristics of esophageal speech. *Journal of Speech and Hearing Research, 19,* 678–689.

Damste, P. H. (1958). *Oesophageal speech after laryngectomy*. Groningen, Netherlands: Hoitsema.

Delattre, P., Liberman, A. M., Cooper, F. S., & Gerstman, F. J. (1952). An experimental study of the acoustic determinants of vowel color: Observations on one- and two-formant vowels synthesized from spectrographic patterns. *Word, 8,* 195–210.

Di Benedetto, M.-G. (1989). Frequency and time variations of the first formant: Properties relevant to the perception of vowel height. *Journal of the Acoustical Society of America, 86,* 67–77.

Diedrich, W. M., & Youngstrom, K. A. (1966). *Alaryngeal speech*. Springfield, IL: Charles C. Thomas.

Fahey, R. P., & López-Bascuas, L. E. (1994). The perception of vowel height in Castillian Spanish: Effects of varying F1-F0 Bark distance. *Journal of the Acoustical Society of America, 96,* 3283 (A).

Flege, J. E. (1991). The interlingual identification of Spanish and English vowels: Orthographic evidence. *Quarterly Journal of Experimental Psychology, 43,* 701–731.

Fox, R. A., Flege, J. E., & Munro, M. J. (1995). The perception of English listeners: A multidimensional scaling analysis. *Journal of the Acoustical Society of America, 97,* 2540–2551.

Gimenez, F., Marco-Algarra, J., Cano, B., Mallea, I., Morant, A., & Orts, M. (1993). Estudio comparativo a largo plazo (cinco años) de pacientes con prótesis fonatoria: Miotomía versus no-miotomía [Long time comparative study of patients with phonatory prostheses: Myotomy versus non-myotomy]. *Acta Otorinolaringológica Española, 44,* 111–116.

Golderos, A. (1984). Modelos para reconocimiento de palabras con independencia del locutor [Speaker-independent word recognition models]. *Mundo Electrónico, 144,* 81–88.

Herrmann, I. (1978). Valve-neoglotis after laryngectomy by cartilage implantation into the trachea. *Laryngology, Rhynology and Otology, 57*, 489–493.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America, 97,* 3099–3111.

Hillenbrand, J., & Nearey, T. M. (1999). Identification of resynthesized /hVd/ utterances: Effects of formant contour. *Journal of the Acoustical Society of America, 105*, 3509–3523.

Jones, D. (1960). *An outline of English phonetics.* Cambridge: W. Heffer & Sons**.**

Klein, W., Plomp, R., & Pols, L. C. H. (1970). Vowel spectra, vowel spaces, and vowel identification. *Journal of the Acoustical Society of America, 48,* 999–1009.

Kytta, J. (1964). Finnish oesophageal speech after laryngectomy: Sound spectrographic and cineradiographic studies. *Acta Otolaryngologica (Stockholm), 195,* 7–91.

Ladefoged, P. (1967). *Three areas of experimental phonetics.* London: Oxford University Press.

Lehiste, I., & Peterson, G. E. (1961). Transitions, glides, and diphthongs. *Journal of the Acoustical Society of America, 33,* 268–277.

Lindblom, B. (1986). Phonetic universals in vowel systems. In J. Ohala & L. Jaeger (Eds.), *Experimental phonology* (pp. 13–44). New York: Academic Press.

Miralles, J. L., & Cervera, T. (1995). Voice intelligibility in patients who have undergone laryngectomies. *Journal of Speech and Hearing Research, 38,* 564–571.

Moon, J. B., & Weinberg, B. (1987). Aerodynamic and myoelastic contributions to traqueoesophageal voice production. *Journal of Speech and Hearing Research, 30,* 387–395.

Onslow, M., van Doorn, J., & Newman, D. (1992). Variability of the acoustic segment durations after prolonged-speech treatment for stuttering. *Journal of Speech and Hearing Research, 35,* 529–536.

Perelló, J., & Más-Dalmau, J. (1980). *Exploración fonoaudiológica* [Fonoaudiology exploration]. Barcelona: Científico-Médica.

Peterson, G., & Barney, H. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America, 24,* 175–184.

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America, 32,* 693–703.

Quilis, A. (1981). *Fonética acústica de la lengua española* [Acoustical phonetics of Spanish language]. Madrid: Gredos.

Rabiner, L. R., & Shaffer, R. W. (1978). *Digital processing of speech signals*. Engelwood Cliffs, NJ: Prentice Hall.

Robbins, J., Christensen, J., & Kemspter, G. (1986). Characteristics of speech production after traqueoesophageal puncture: Voice onset time and vowel duration. *Journal of Speech and Hearing Research, 29,* 499–504.

Robbins, J., Fisher, H. B., Blom, E. D., & Singer, M. I. (1984). A comparative acoustic study of normal, esophageal, and tracheoesophageal speech production. *Journal of Speech and Hearing Disorders, 49,* 202–210.

Rosner, B. S., & Pickering, J. B. (1994). *Vowel perception and production.* Oxford: Oxford University Press

Sisty, M., & Weinberg, B. (1972). Formant frequency characteristics of esophageal speech. *Journal of Speech and Hearing Research, 15,* 439–448.

Smith, B. L., Hillenbrand, J., & Ingrisano, D. R. (1986). Comparison of temporal measures of speech using spectrograms and digital oscillograms. *Journal of Speech and Hearing Research, 29,* 270–274.

Stevens, K. N., & House, A. S. (1955). Development of a quantitative description of vowel articulation. *Journal of the Acoustical Society of America, 27,* 484–493.

Swanson, L. A., Leonard, L. B., & Gandour, J. (1992). Vowel duration in mother's speech to young children. *Journal of Speech and Hearing Research, 35,* 617–625.

Turner, G. S., Tjaden, K., & Weismer, G. (1995). The influence of speaking rate on vowel space and speech intelligibility for individuals with amyotrophic lateral sclerosis. *Journal of Speech and Hearing Research, 38,* 1001–1013.

Received July 24, 2000

Accepted April 25, 2001

DOI: 10.1044/1092-4388(2001/077)

Contact author: Teresa Cervera, PhD, Universitat de València, Department ode Psicologia Basica, Avda, Blasco Ilanez, 21, 46010 València, Spain.
E-mail: Teresa.Cervera@uv.es

Appendix. List of words in the acoustical analysis.

Pino	/pino/	(pine)	**Zona**	/θona/	(zone)
Letra	/letra/	(letter)	**Templo**	/templo/	(church)
Musa	/musa/	(muse)	**Fase**	/fase/	(phase)
Ganso	/ganso/	(gander)	**Perla**	/perla/	(pearl)
Cero	/θero/	(zero)	**Goma**	/goma/	(rubber)
Milla	/miλa/	(mile)	**Nube**	/nube/	(cloud)
Vega	/bega/	(fertile plain)	**Rosa**	/ŕosa/	(rose)
Tira	/tira/	(strip)	**Dique**	/dike/	(dike)
Seda	/seda/	(silk)	**Gasa**	/gasa/	(gauze)
Rima	/ŕima/	(rhyme)	**Jaspe**	/xaspe/	(jasper)
Poda	/poda/	(pruning)	**Cita**	/θita/	(appointment)
Liso	/liso/	(plain)	**Chino**	/ʧino/	(Chinese)