Borrower: GDC

Lending String: *ORC,MTG,MZF,XFF,WUG

Patron: Amith, Jonathan

Journal Title: American Indian linguistics and ethnography in honor of Laurence C. Thompson /

Volume: **Issue:**
Month/Year: 1993 **Pages:** ? no information

Article Author:

Article Title: Kirk, P. L., J. Ladefoged, and P. Ladefoged; Quantifying acoustic properties of modal, breathy and creky vowels in Jalapa Mazatec.

Imprint: Missoula, MT ; University of Montana , 1

ILL Number: 69879282

ARIEL
ODYSSEY ENABLED

Charge
Maxcost: $25IFM

Shipping Address:
Gettysburg College Library-ILL
300 N. Washington St.
Gettysburg, PA 17325-1493

Fax:
Ariel:
Odyssey: 206.107.42.97

QUANTIFYING ACOUSTIC PROPERTIES OF MODAL, BREATHY, AND CREAKY VOWELS IN JALAPA MAZATEC

PAUL L. KIRK, JENNY LADEFOGED, and PETER LADEFOGED

This paper[1] is intended to be both a detailed account of an interesting phenomenon in American Indian languages, and also a description of new methods of quantifying phonetic data. We hope that it will be of interest to both Americanists such as Larry Thompson, who has contributed significantly to the detailed study of American Indian languages, as well as to a more general audience of phoneticians and linguists. The specific problem we will discuss is the nature of the vocal cord activity in certain phonological contrasts in Jalapa Mazatec.

Obviously the best way of evaluating what the vocal cords are doing is to look at them. If we want to quantify different types of phonation, the most direct technique is to measure the glottal movements observed by means of high speed cinematography, and the muscular actions as recorded by electromyography. But linguists frequently have at their disposal only an ordinary tape recording, often one made in the field rather than under ideal laboratory conditions. In these circumstances they can usually investigate different types of phonation simply by reference to the kinds of analysis provided by a sound spectrograph. It is therefore important to evaluate the ways in which this instrument can be used for measuring different types of phonation recorded in field conditions. Thus, in this paper we analyze, with the Kay Digital Spectrograph, the three different voice qualities of Jalapa Mazatec vowels.

---

[1] A version of this paper was presented at the 106th Meeting of the Acoustical Society of America (November 7-11, 1983 and appeared in UCLA working papers in phonetics 59 (March 1984). We gratefully acknowledge the support for our work of USPHS grant NS 18163-02. Also Terence and Judith Schram provided assistance in the field during the time that the data for this paper was collected and we wish to thank them for clarifying and deepening our understanding of Mazatec through many happy hours of discussion.

The term voice quality is used in a variety of ways. Laver (1980) uses it to describe any long term characteristic of a speaker, including things such as persistent pharyngealization or labialization. We will use the term voice quality in a more restricted way, applying it only to those aspects of speech that are due to the action of the vocal cords. Differences of phonation type of this kind are sometimes referred to as breathy voice, creaky voice, laryngealized voice, etc.

The major problem with these labels is that they are used in different ways by different people. Furthermore, they are often used to identify purely individual characteristics of voices. Phoneticians need to develop a set of terms that could be used to describe not just one voice at a time, but properties shared systematically by a whole group of voices. These requirements may be satisfied by using material from languages which distinguish meanings by changes in voice quality.

Most languages distinguish between just voiced and voiceless sounds, and it does not matter whether the voice sounds breathy or harsh. If one speaks with very loose vibrations of the vocal cords, allowing a greater airflow than normal, a laryngologist might prescribe therapy; but such a voice quality does not affect the linguistic contrasts between words. There are, however, a number of American Indian and other languages in which differences in the mode of vibration of the vocal cords reflect differences in meaning. When a speaker of Mixtec or Navaho uses a change in voice quality to produce a change in meaning, then everyone who understands the language has to be able to recognize this change. Moreover all speakers have to make the change in much the same way. These languages therefore offer material that can be used for evaluating methods of measuring phonation types because one knows, *a priori*, that the sounds will differ in a given way. Individual speakers may have their own idiosyncratic characteristics; but for each speaker, if the meaning requires a word with a breathy voiced vowel in one instance, a creaky voiced vowel in another, and a modal voiced vowel in still another, then the measurable difference between the three words is a measurable difference between breathy voice, creaky voice, and modal voice. It is still, of course, possible that what we call creaky voice in Navaho may not be the same as what we call creaky voice in Mixtec. However, if we find that the measures we use for quantifying the difference in one language are also reliable indicators

of the difference in other languages, then we can conclude that we are in fact measuring difference between phonation types that can be defined. Although in this paper we have not attempted any cross-linguistic comparisons among American Indian languages, we hope that our methodology will have put the study of differences in phonation type on a firmer basis, so that such studies can be better conducted in the future.

There is a cost to using real language material that does not apply to artificially produced samples of different phonation types. Many clinical approaches to the study of voice quality (e.g. Davis 1976) use samples elicited by asking the patient to produce a steady state vowel lasting for several seconds. In real languages the differences occur in running speech; consequently they may be difficult to analyze because they may last for only a small part of a second. But this cost is more than offset by the advantage of being able to use several speakers, knowing that they must all be behaving in some similar ways, if they are producing words which their listeners can identify correctly.

LANGUAGE DATA. The language we chose to examine is the Jalapa de Diaz dialect of Mazatec (hereafter referred to as Jalapa Mazatec), spoken by approximately 8,000 people in the District of Tuxtepec, Oaxaca, Mexico. Twenty three dialects of Mazatec have been identified (Kirk 1970) and a number of similarities as well as differences among the dialects are clearly observable (Kirk 1966). Mazatec is a member of the Popolocan branch (Gudschinsky 1959) of Otomanguean (Rensch 1976). The syllable structure of Huautla de Jiménez Mazatec, described by Pike and Pike (1947), has many similarities to other dialects of Mazatec (Pike 1956, Kirk 1966, Jamieson 1977a, Jamieson 1977b, Schram and Pike 1978).

Among the various dialects of Mazatec we chose Jalapa Mazatec because it is rich in voice quality distinction. In fact it probably makes a greater use of differences in phonation types than any other language we know of. In consonants, it makes use of voiced and voiceless oppositions as well as aspirated and unaspirated oppositions. In addition Jalapa Mazatec makes use of modal voice, creaky voice, and breathy voice for lexical differentiation. It distinguishes words such as 'horse' and 'arse' by the former having breathy voice and the latter having creaky (all other features of these words are the same, i.e., [ndæ̤¹] 'horse', [ndæ̰¹] 'arse').

Jalapa Mazatec is a tone language contrasting three pitch levels and glides between them. Contrastive tones differentiate meanings of words as well as grammatical categories and, along with nasalization, are the domain of the syllabic nucleus (Pike and Pike 1947). In this paper we write low tone with the widely used convention of superscript one. However, most publications on Otomanguean languages use the superscript numbers in reverse fashion with one representing high tone and increasing larger number representing lower tones.

In addition to the distinction between breathy voice and creaky voice, there is an opposition between modal voice and breathy voice as in:

[ni^{2}mæ^2]	'bumblebee'	[mæ̤^{2}]	'he wants'
[nda^2]	'good'	[nda̤^{12}]	'hard'
[ni^{2}ʔja^2]	'house'	[ja̤^{2}]	'he wears'
[jæ^1]	'snake'	[tʃu^{1}jæ̤^{1}]	'turtle'
[ni^2]	'red'	[ja^{3}ni̤^{2}]	'he carries on his back'
[ju^2]	'willing'	[ti^{3}jṳ^{2}]	'it is stopped'

There is also an opposition between modal voice and creaky voice as in:

[ʃĩ^1]	'hunt'	[ʃḭ̃^{1}]	'male'
[nũ^3]	'year'	[nṵ̃^{3}]	'vine'
[kã^2]	'twenty'	[kã̰^{2}]	'alone'
[tʃa^2]	'old'	[tʃa̰^{2}]	'load'
[tʰæ^2]	'itch'	[tʰæ̰^{2}]	'sorcery'
[hi^{3}tsi^1]	'yours'	[tsḭ^1]	'nausea'

Three way opposition, though not minimal between modal voice, creaky voice, and breathy voice, is attested by the following examples:

[ja^3] 'tree', [ni^{2}ʔja^2] 'house' [ja̰^{2}] 'he carries' [ja̤^{2}] 'he wears'
[ntʰæ^1] 'seed' [ndæ̰^{1}] 'arse' [ndæ̤^{1}] 'horse'

Although a detailed examination of the phonology of Jalapa Mazatec is not the focus of this paper, nevertheless from the discussion thus far, it is evident that the phonology is fairly complex. Examples cited above

show that differences in phonation type are not just an incidental phenomenon associated with a few words, but a pervasive characteristic of the language. If phonation type is taken to be a property of a syllable, there is a three way contrast between modal voice, creaky voice, and breathy voice. If it is regarded as a property of a segment, then we may say that there is an opposition between breathy voiced and modal voiced consonants, and between creaky and modal voiced vowels.

The material we used in our analysis comes from a recording of a longer word list illustrating a wide range of phonological properties. There were five speakers, all of whom spoke the same dialect; in December 1982 four were recorded in Jalapa de Diaz and the fifth was recorded in Mexico City. For the purposes of the present analysis we concentrated on three words which were carefully matched for pitch and vowel quality. Two of the words, [ndæ̰^{1}] 'arse' and [ndæ̤^{1}] 'horse', are differentiated only by creaky voice as opposed to breathy voice. The third word used in this study [ntʰæ^1] 'seed' has a vowel with modal voice of the same quality and tone as the vowels with creaky and breathy voice.

WIDE BAND SPECTROGRAMS. We produced a number of different displays of this data. Figure 1 shows frequency-amplitude-time displays, using a 300 Hz bandwidth filter (wide band spectrograms) illustrating vowels with creaky voice, modal voice, and breathy voice for all five speakers. These vowels have been segmented from the words [ndæ̰^{1}] 'arse' (creaky voice), [ntʰæ^1] 'seed' (modal voice), and [ndæ̤^{1}] 'horse' (breathy voice). Wide band spectrograms provide good displays of what Fant (1973) calls the F-pattern, the overall pattern of the formants during a sound. The formants are particularly clear during the creaky voice vowels in the top row, and fairly evident during the modal voice in the second row.

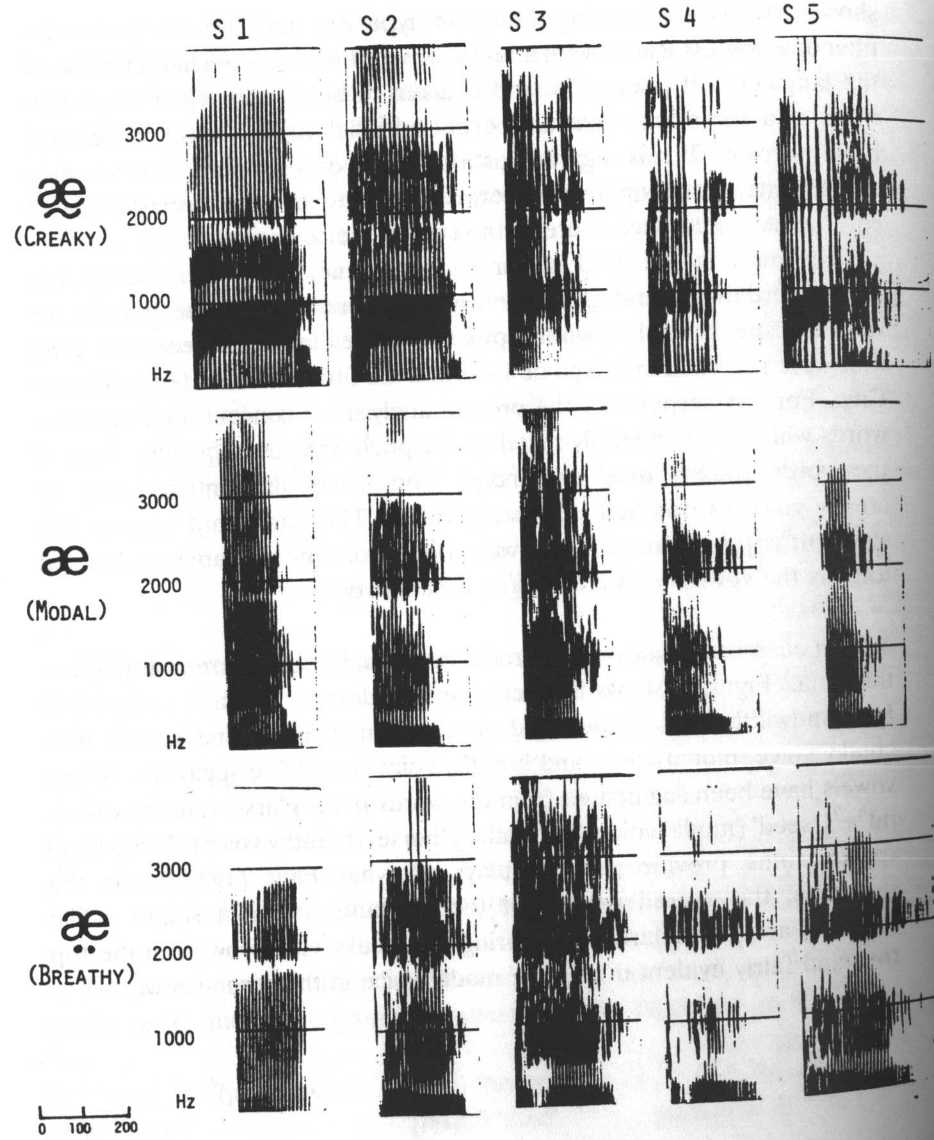

Figure 1. Wide band spectrograms of Jalapa Mazatec creaky, modal, and breathy vowels for five speakers, SI-S5.

In the breathy vowels in the bottom row the first formant is less well defined. In general the formants are in similar places in all three phonation types. There may be a tendency for the first formant to have a slightly higher frequency during creaky vowels—a difference that would be associated with the raising of the larynx and the consequent shortening of the vowel tract during creaky phonation. We were, however, unable to use these displays to provide reliable measures of formant frequencies that could be used to quantify phonation types.

Wide band spectrograms also provide a very suitable basis for duration measurements. The displays of the vowels in Figure 1 were carefully segmented from the spectrograms of the complete utterances. They begin with the first pulse of the vocal cords after the release of the consonant. It is arguable that this may not be the beginning of the vowel, as the spectrograms show the formants associated with the movements of the tongue away from the alveolar ridge. These are particularly noticeable in the case of the creaky vowels in the first row. The modal vowels follow an aspirated [tʰ], and the transitions are over before voicing starts. The breathy vowels in the bottom row do not have such a clear first formant, but nevertheless do exhibit some evidence of consonant transitions. Despite the presence of transitions, we took the first pulse of the vocal cords as the start of the vowel because it enables us to make more reliable measures; we could not find a satisfactory procedure for defining the end of the transition.

Defining the end of the vowel was also difficult, as there is no following consonant (Mazatec has no closed syllables). For the purposes of this study it was taken to be the last point at which there is energy in at least two of the formants. Given these definitions it is quite clear that both the creaky and breathy vowels are longer than the modal vowels. The mean durations for the five speakers are: creaky vowels 266 msec (s.d. 58); modal vowels 174 msec (s.d. 9); breathy vowels 244 msec (s.d. 49). Some of this difference may be associated with the differences in the contexts: the modal vowels followed an aspirated consonant cluster [ntʰ], whereas the creaky and breathy vowels followed a voiced cluster [nd]. But our observations of the length differences in the other contrasts reported above also provide support for the conclusion that the breathy and creaky phonation types are regularly accompanied by greater duration. For example, the mean duration of the breathy vowel in [mæ̤] 'he wants' is 317 msec (s.d.40), while that of the modal vowels in [ni²mæ²] 'bumblebee' is 177 msec (s.d. 27).

A number of points that are more easily quantified with the aid of other types of displays are clearly visible in the general picture provided by the wide band spectrograms in Figure 1. Most importantly it may be seen that the vowels in the first row with creaky voice are marked by jitter. The vertical striations (i.e. glottal pulses) occur at irregularly spaced intervals. For all speakers, the pattern is for pulses to be grouped closer together at the onset of the vowel followed by increased distances between pulses moving toward the center of the vowel, followed by decreased distances between pulses toward the coda of the vowel.

Despite the slower rate of the glottal pulses during part of this vowel, this word is perceived by Mazatec speakers as having the same lexical tone as the modal vowel word. Mazatec speakers whistle these words with the same pitch in whistle speech. Surprisingly, phoneticians also perceive the pitch of these creaky and modal vowels as being the same. The laryngealization superimposed in the middle of the word does not offset the auditory impression created by the more regular voicing at the beginning and end of these vowels.

Qualitative differences in spectral balance are evident in Figure 1. The higher frequencies tend to be more clearly visible during the creaky vowels (see especially speakers two, four, and five). Yet another noticeable difference between the vowels in the second and third row compared with the vowels in the first row is that the bandwidth of each formant is somewhat less in the vowels with creaky voice (row one). It is also clear that the overall differences between the vowels in breathy voice (row three) in comparison with those in modal voice (row two) are similar, but in the opposite direction, to the differences that occur between creaky voice and modal voice. Furthermore, we may note that for all speakers, breathy voice is more clearly seen in the onset part of the vowel since the coda section of the vowel tends to have modal voice. Finally, we can see that for two of the five speakers, the amplitude of the first formant is distinctly less in the breathy vowels. Thus the wideband spectrograms provide an excellent general view of many differences in phonation types, some of which can be quantified by reference to other kinds of display.

POWER SPECTRA. Another kind of display that can be produced with the aid of a sound spectrograph is a power spectrum showing the relative intensities of the component frequencies. Figure 2 is a display of power spectra of the three phonation types. For these spectra, a 45 Hz (narrow band) filter was used so that the amplitudes of each of the component harmonics are clearly visible.

We noted in the discussion of the wide band spectrograms that the higher formants are more evident during creaky voice. Narrow band power spectra offer a way of quantifying this spectral tilt. But before we consider a method of measuring the relative spectral balance associated with each phonation type, we must note that there are two different ways in which higher frequencies can come to have more energy. When producing breathy voice the vocal cords are vibrating more loosely, often not making complete contact along their whole lengths at any time in the glottal cycle. As a result there is a greater rate of airflow through the glottis producing a turbulent airflow with more random high frequency components. When producing creaky voice the vocal cords are much more tensed, closing rapidly during each glottal cycle. As a result the vocal tract is excited by a sharper pulse that has more energy in the higher harmonics. In order to separate out these different phonation types we need a measure that distinguishes whether the increase in the higher frequencies is associated with additional components produced by a semi-random, turbulent, airflow, or by an increase in the intensity of the higher harmonics produced by a sharper glottal pulse. None of the displays produced by the spectrograph provides a way of separating out the turbulent airflow so that it can be quantified. But the power spectra enable us to quantify the relative amount of energy in different harmonics. We chose as our measure the difference in dB between the intensity of the fundamental and the intensity of the largest harmonic in the first formant.[2] This measure can be used for comparing phonation types only

[2]We are grateful to Professor K. N. Stevens for suggesting this measure to us.

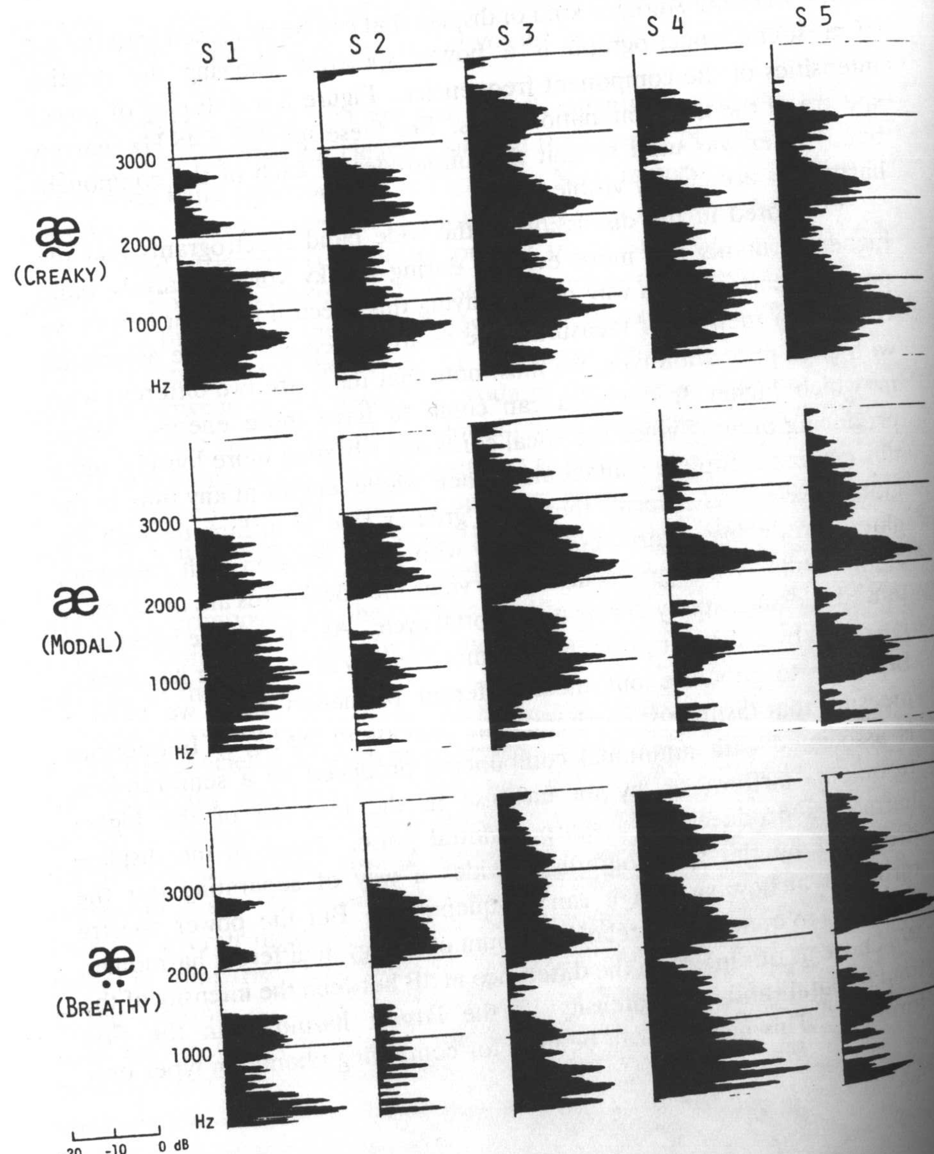

Figure 2. Spectra of Jalapa Mazatec creaky, modal, and breathy vowels.

in cases in which the vowels being compared have similar formant frequencies, as the relative intensity of each formant is a function of its frequency (Fant 1956). In the case of these Jalapa Mazatec vowels this is perfectly appropriate.

The difference between the amplitude of the fundamental and that of first formant is displayed in figure 3. For the five speakers, the mean for creaky voice (indicated in the bar graph with diagonal striations) is -17 dB with a standard deviation of 3.7 (i.e., the fundamental has 17 dB less amplitude than the first formant). The mean for modal voice (indicated by the black bar) is -6.6 dB with a standard deviation of 4.4. The mean for breathy voice (indicated by a cork-screw pattern) is +5.2 dB with a standard deviation of 3.8. Note that there is considerable variation from speaker to speaker in the three phonation types; but for each speaker on this measure the value for creaky voice is less than that for modal voice, and the value for modal voice is less than that for breathy voice. This measure separates out breathy voice successfully in that for all speakers the value for breathy voice is higher than that for modal voice for any speaker. But it is less successful in distinguishing between modal voice and creaky voice in terms of their absolute values; on this measure speaker three's modal voice has an absolute value less than speaker four's creaky voice. It is only in relative terms that each speaker markedly differentiates the three phonation types on this measure.

Waveforms. Yet another useful display provided by the digital sound spectrograph is that of the waveform. As noted earlier, the wide band spectrograms show that creaky vowels have speech jitter (i.e. irregularly spaced pulses). In the waveform displays these irregular intervals are more clearly evident than in the wide band spectrograms.

Displayed in Figure 4 are the first 105 msec of the creaky vowels for all speakers, and for comparison the modal and breathy vowel displays for speaker five. The breathy vowel is characterized by an onset of indiscernible pulses; the modal vowel has regular pulses. Our measure of the jitter is the variation in the interval between adjacent pulses. It can be seen in these displays that the onset of each creaky vowel pulse is clearly discernible for each of the speakers except for speaker three. For this

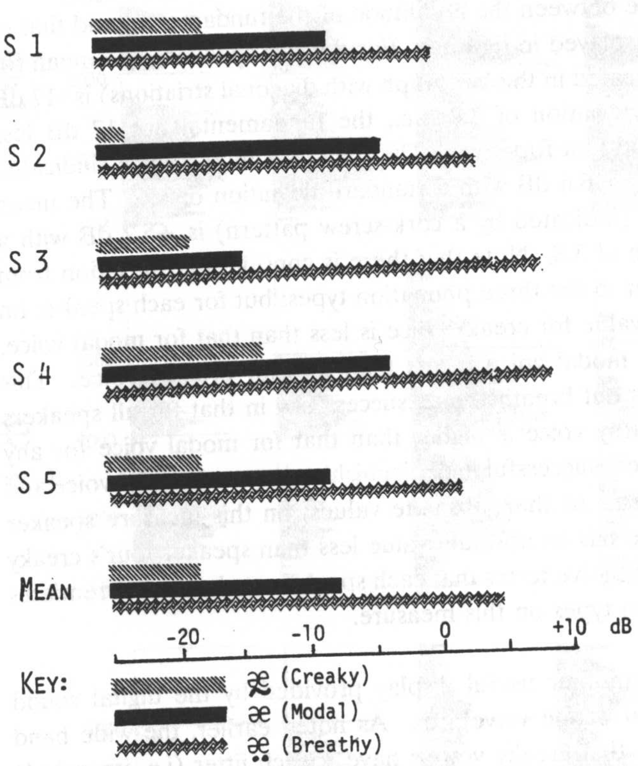

Figure 3. Relationship of the fundamental to the first formant in Jalapa Mazatec vowels.

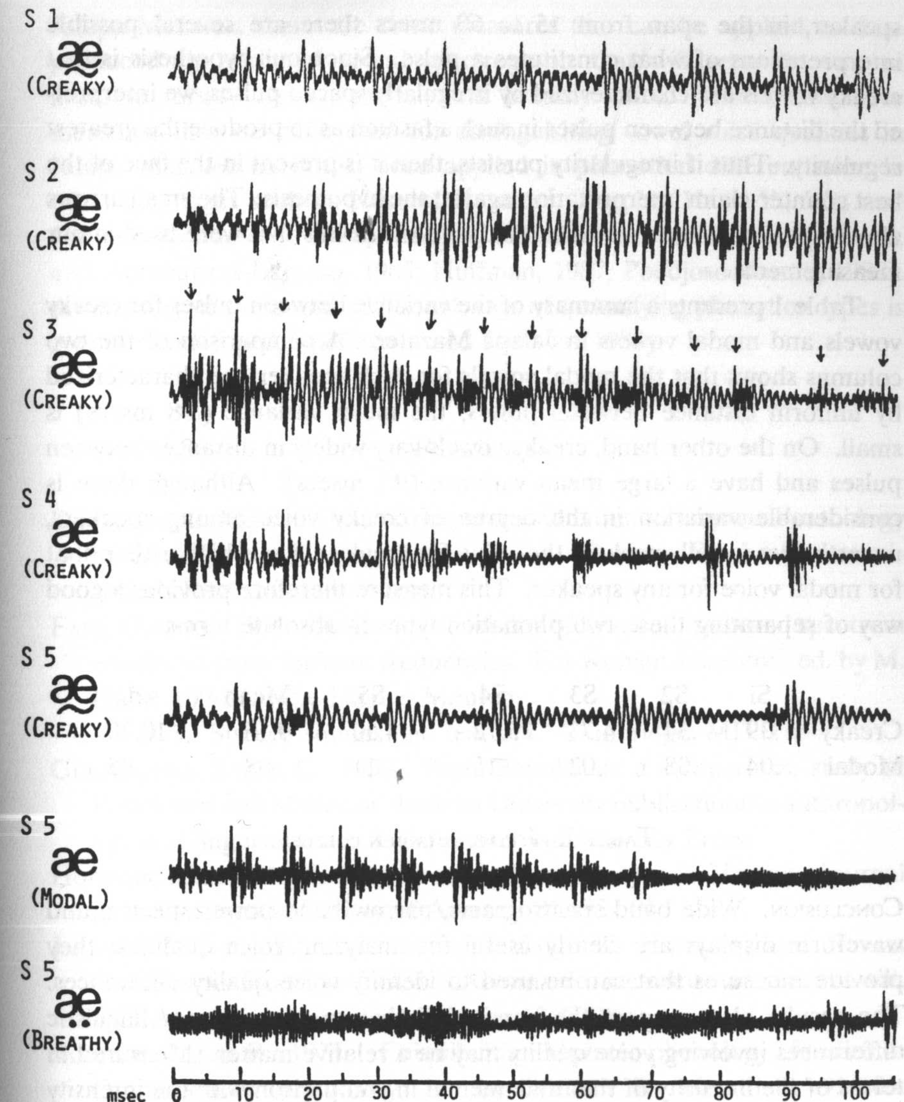

Figure 4. Waveforms of Jalapa Mazatec creaky vowels for five speakers and modal and breathy vowels for the fifth speaker.

speaker, in the span from 15 to 60 msecs there are several possible interpretations of what constitutes a pulse. Since our hypothesis is that creaky vowels are characterized by irregularly spaced pulses, we interpreted the distance between pulses in such a fashion as to produce the greatest regularity. Thus if irregularity persists, then it is present in the face of the best counter-claim interpretation against the hypothesis. The small arrows above speaker three's waveform indicate the points that were used in our measurements.

Table 1 presents a summary of the variance between pulses for creaky vowels and modal vowels in Jalapa Mazatec. A comparison of the two columns shows that the modal vowels for each speaker are characterized by uniform distance between pulses; the mean variance (.08 msecs) is small. On the other hand, creaky vowels vary widely in distances between pulses and have a large mean variance (9.1 msecs). Although there is considerable variation in the degree of creaky voice among speakers, nevertheless for all speakers the value for creaky voice is higher than that for modal voice for any speaker. This measure therefore provides a good way of separating these two phonation types in absolute terms.

	Sl	S2	S3	S4	S5	Mean	s.d.
Creaky	.39	.34	4.73	25.72	14.50	9.10	10.90
Modal	.04	.08	.02	.15	.11	.08	.05

Table 1. Jitter between pulses

Conclusion. Wide band spectrograms, narrow band power spectra, and waveform displays are clearly useful for analyzing voice qualities; they provide measures that can be used to identify voice quality differences. The results also support the conclusion that to some extent linguistic differences involving voice quality may be a relative matter. Measured in terms of the intensity of the fundamental in comparison with the intensity of the first formant, what may be modal voice for one speaker may count as creaky voice for another. But in this respect voice quality is similar to other phonetic features such as vowel height: measured in terms of frequency of the first formant, what may count as a high vowel [ɩ] for some speakers may count as a mid vowel [ɛ] for others. Furthermore, in

Jalapa Mazatec, there are some measures that can be applied to data produced by a digital sound spectrograph that will separate out, in absolute terms, creaky, modal, and breathy voice. The quantitative measures that we have used for distinguishing vowel voice qualities in Jalapa Mazatec have also been applied in phonation differentiation in other languages such as Hmong, Burmese, Jingpho, Hani, Yi, Wa, and Javanese (Maddieson and Ladefoged, 1985; Jackson, Ladefoged, Huffman, and Antoñanzas-Barroso, 1985; Huffman, 1987; Poedjosoedarmo, 1988). A summary of these and additional ways of quantifying phonation types is given in Ladefoged, Maddieson and Jackson (1988).

References

Davis, Steven Bruce. 1976. Computer evaluation of laryngeal pathology based on inverse filtering of speech. Santa Barbara: Speech Communications Research Laboratory.

Fant, Gunnar. 1956. On the predictability of formant levels and spectrum envelopes from formant frequencies. For Roman Jakobson, ed. by M. Halle, 109-20. The Hague: Mouton.

——. 1973. Speech sounds and features. Cambridge: MIT Press.

Gudschinsky, Sarah C. 1959. Proto-Popolocan: a comparative study of Popolocan and Mixtecan. Indiana University publications in anthropology and linguistics, no. 5. Baltimore, Md.:Waverly Press.

Huffman, Marie. 1987. Measures of phonation type in Hmong. Journal of the Acoustical Society of America 81.495-504.

Jackson, Michel, Peter Ladefoged, Marie K. Huffman, and Norma Antoñanzas-Barroso. 1985. Automated measures of spectral tilt. UCLA working papers in phonetics 62.77-88.

Jamieson, Allan R. 1977a. Chiquihuitlan Mazatec phonology. Studies in Otomanguean phonology, ed. by William R. Merrifield, 93-105. Arlington: Summer Institute of Linguistics, University of Texas.

——. 1977b. Chiquihuitlan Mazatec tone. Studies in Otomanguean phonology, ed. by William R. Merrifield, 107-35. Arlington:Summer Institute of Linguistics, University of Texas.

KIRK, PAUL L. 1966. Proto-Mazatec phonology. Ph.D. dissertation, University of Washington.

——. 1970. Dialect intelligibility testing: the Mazatec study. IJAL 36.205-11.

LADEFOGED, PETER, IAN MADDIESON and MICHEL JACKSON. 1988. Investigating phonation types in different languages. Voice physiology: voice production, mechanisms and functions, ed. by Osamu Fujimura. New York: Raven Press.

LAVER, JOHN. 1980. The phonetic description of voice quality. Cambridge: Cambridge University Press.

MADDIESON, IAN and PETER LADEFOGED. 1985. Tense and lax in four minority languages of China. Journal of phonetics 13.433-54.

PIKE, EUNICE V. 1956. Tonally differentiated allomorphs in Soyaltepec Mazatec. IJAL 22.57-71.

PIKE, KENNETH L. and EUNICE VICTORIA PIKE. 1947. Immediate constituents of Mazateco syllables. IJAL 13.78-91.

POEDJOSOEDARMO, GLORIA R. 1988. A phonetic description of voice quality in Javanese traditional female vocalists. Asian music 19.2.93-126.

RENSCH, CALVIN R. 1976. Comparative Otomanguean phonology. Language science monograph 14. Bloomington:Indiana University Press.

SCHRAM, JUDITH L. and EUNICE V. PIKE. 1978. Vowel fusion in Mazatec of Jalapa de Diaz. IJAL 44.257-61.