VOWEL ACOUSTICS IN AKUNTSÚ: DISPERSION AND NON-MODAL PHONATION

Fernando O. de Carvalho[1]
Carolina Coelho Aragon[2]

fernaoorphao@gmail.com
carolina.aragon@utah.edu

ABSTRACT: This paper aims to present a preliminary analysis of creaky or laryngealized vowels in Akuntsú spoken by only six monolingual indians on the Omerê River, at the Brazilian western state of Rondônia, Brazil. Vowel spaces are re-expressed in terms of a standard psychoacoustic scale and further quantitative information is addressed. A global measure of acoustic dispersion is presented and the value obtained for the Akuntsú vowel space is compared to similar values obtained from an independent sample of languages. Results suggest that non-modal phonation in Akuntsú can be characterized as the effect of at least two distinct articulatory mechanisms and their differing acoustic effects.
KEYWORDS: Akuntsú language; Non-modal phonation; Vowel space; Acoustics.

INTRODUCTION

Akuntsú is a member of the Tuparí family of the Tupían stock. This language is among the most highly endangered languages in the Americas, due to its reduced number of speakers and due to the fact that no younger generation is learning Akuntsú as its native language. The remaining six monolingual survivors of a massacre, to which the Akuntsú ethnic group was subjected, were first contacted by *Fundação Nacional do Índio* (National Foundation of Indians) in 1995. After that contact, they were free to get along with their traditional life-style in a small part of the Omerê River surrounding region, within the state of Rondônia. Given their recent contact, the work of linguistic documentation began only in 2004.

[1] *Laboratório de Línguas Indígenas* (LALI), Universidade de Brasília.
[2] *Center for American Indian Languages* (CAIL), University of Utah. *Laboratório de Línguas Indígenas* (LALI), Universidade de Brasília.

This paper aims to improve the knowledge regarding the sound structure of this endangered language, presenting a perceptually grounded measure of global dispersion for the language's vowel space, whose value is compared to those computed for 28 other languages from an independent sample. The second part of the paper provides a brief description of the phenomenon of creaky or laryngealized phonation (a kind of non-modal phonation) in Akuntsú, a very conspicuous phonetic trait of this language.

The paper is organized as follows: section 1 provides a brief background of the phonological and phonetic system of Akuntsú language; section 2 presents the methodology of analysis and the procedures of data collection; section 3 addresses the results and discusses their implications; then section 4, provides a conclusion and future research.

1. Background on the Language

The Akuntsú consonantal inventory is formed by 11 independent segments - /p, t, k, ʔ, tʃ, m, n, ŋ, r, w, j/ - patterned in terms of 5 places of articulation – bilabial, alveolar, alveolar-palatal, velar and glottal – as well as 4 manners – stop, affricate, nasal and approximant. Given the concern of the present paper with the acoustic expression of the language's vowels, the following sections do not deal in further detail with the phonetic implementation of consonantal contrasts (for details see Aragon 2008).

1.1 The vowel inventory of Akuntsú

The phonetic realizations and the phonemic inventory of the oral vowels in Akuntsú are described below:

	Front	Central		Back
High	[i] [ĩ]	[ɨ] [ɨ̰] [ɨ̃]	[ɯ] [ɯ̃]	[u] [ũ] [ṵ] [ṵ̃]
	[ɪ]			[ʊ]
Mid	[e]			[o] [õ] [o̰] [õ̰]
	[ɛ] [ɛ̃] [ɛ̰] [ɛ̰̃]	[ɐ]	[ʌ]	[ɔ] [ɔ̰]
Low	[æ]	[a] [ã] [a̰] [ã̰]		

Table 1: Phonetic Chart with some of the vowel qualities observed in the systematic phonetic representations of Akuntsú.

	Front	Central	Back
[+high]	i	ɨ	o
[-high]	e	a	

Table 2: Chart showing the five vowel phonemes in Akuntsú.

The inventory of Akuntsú shows three high and two low phonemic vowels. However only the front and central vowels differ as a function of the [+/- high] feature specification. There are also four phonemic nasals vowels /ã, ẽ, ĩ, õ/, however they are not the focus of the present study.

Akuntsú also has surface creaky-voiced vowels which are not phonologically contrastive. Aragon (2008) shows that oral or nasal vowels become laryngealized when preceded by a glottal stop in stressed syllable position, as shown in (1) below:

(1)

[peˈʔa̰] ‘firewood’ [kwaˈʔḭ] ‘stone’

[ˈʔṵ] ‘tongue’ [ɛˈʔɨ̰] ‘blood’

However, laryngealized vowels are also found in other environments as follows:

(2)

[ḭˈpɛ̰bṵ] ‘its feather’ [ɔ̰ɾɔ̰ˈdḭta̰] ‘sloth (sp.)’

[ɔ̰ɾɔ̰ˈgʷḭ]   ‘peanut’   [kɨ̰ˈbɛ̰g˺] ‘papaya’

Even though the presence of non-modal, laryngealized vowels may be related to the phonetic implementation of stress (as discussed in section 4 below) some paralinguistic factors may also play a role in conditioning its pattern of occurrence. In the narratives of past traumatic or stressful events, for instance, one often finds laryngealization or creaky voicing spread all over the syllables of a word (cf. Aragon 2008). That laryngealization is optionally spread over preceding syllables, as seen from fluctuations, such as:

(3)

[ɔ̰ɾɔ̰ˈgʷḭ] ~ [ɔɾɔˈgʷḭ]   ‘peanut’   [kɨ̰ˈbɛ̰g˺] ~ [kɨˈbɛ̰g˺] ‘papaya’

1.2 PREVIOUS DISCUSSION

In a preliminary analysis of the phonetics of Akunstú (Aragon and Carvalho 2007), the first two formants (F1 and F2) of /a, e, i, u, o/ tokens were considered – the high central vowel was excluded due to a lack of a significant number of adequate tokens for this category in our sample data.

Besides providing a preliminary description of vowel phonetics in Akuntsú, this previous work also brought evidences to support the independently motivated hypothesis that a previously postulated /ɔ/ phoneme, originally presented in Cabral and Aragon (2005) encompassed actually a set of allophones of either the phoneme /a/ or the phoneme /u/. It was also shown that, along with arguments of a more phonological or functional sort, there is no reason to consider a low back vowel as part of the vocalic inventory of Akuntsú (see Aragon and Carvalho 2007 for further details).

2. METHODOLOGY

Due to the fact that linguistic fieldwork among the Akuntsú is subject to rather stringent, law-enforced conditions of brief co-existence among a reduced group of individuals, the monolingual survivors of a recent genocide, the data collected are in several senses far from ideal. It is very hard to employ constructed elicitation sets aimed at controlling for all the variables that may have an effect on the dependent measures (besides the independent categorical variables of phonemic vowel category) and the fact that this work is part of the *first* project for the description and documentation of this language contributes to this rather messy situation in which an adequate protocol for the controlling of extraneous variables bearing on the values of our measurements is not feasible

The fieldwork was done by the second author (C.C.A.) in 2004, 2006, 2007 and 2008. Data were recorded on free-field conditions with two adult monolingual speakers of Akuntsú, one male (W) and one female (Y). Data were recorded in digital format (WAV) with a *Sony IC Recorder ICD-MX20*, a *Sony Portable Minidisc Recorder MZ-NH700* and with a *Sony ECM-MS957* stereo microphone with rotating mid capsule. The data subject in the present work was comprised mostly of words pronounced in isolation. The number of tokens for each vowel category, for the male and female speaker were respectively, /a/: (13/11), /i/: (11/11), /e/: (13/11), /o/: (12/7) and /ɨ/: (7/6); total: (56/46). For the analysis of vowel quality by formant peak frequency estimation, no creaky vowels were used, due to the fact that the LPC-based algorithm employed for formant value estimation is based on the assumption that the speech sample analyzed was produced with a glottal source output of a specific spectral shape, one that cannot be assumed for strongly laryngealized speech (Boersma & Weeninck 2006).

Expanding on the preliminary results reported in Aragon and Carvalho (2007), the size or overall dispersion of the Akuntsú vowel space was computed by adding the Euclidean distances between the mean Bark values[3] for each of its three 'point vowels', following the procedure implemented by Livjin (2000). As a first step, formant values in Hz were converted into the psychoacoustic Bark scale using the expressions given in Traunmüller (1990) and presented below for convenience (F stands for formant value in Hz):

$$Z = [26.81\ F / (1960 + F)] - 0.53$$

The formula has an added correction factor for values of $Z < 2.0$:

$$Z' = Z + 0.15\ (2 - Z)$$

[3] The Bark (z) scale provides a perceptually more realistic estimate of the auditory distance between auditory objects originally described in terms of the physical parameter of center frequency (Hz).

Formant estimation (in Hz) was carried with the LPC-based algorithm of the *Praat* program for speech analysis (Boersma and Weeninck 2006) and with the usual aid of visual inspection of wideband spectrograms. Arguments of the algorithm were adjusted for this task as a function of the nature of data (e.g., the argument of the LPC-based algorithm specifying the maximum formant value was lowered to 4500 when analyzing male speech samples, as an optimization procedure to the lowered frequency maximum in male speech when compared to that of a female individual).

Mean values over the distributions of Bark values for each 'point vowel' (/a, i, u/) category were computed, and the bi-dimensional Euclidean distances (i.e., distances between points defined by F1 (Bark), F2 (Bark) values) between these vowels means were estimated[4]**.** The sum of the distance values between pairs of adjacent categories were summed and provided the global measure of dispersion (cf. Livjin 2000). Statistical analysis and the generation of graphs were done on *SPSS 14.0*.

3. Discussion

3.1. Properties of the Akuntsú vowel space

The vowel space of Akuntsú over the mean Z values for the sample under study is shown below in Figure 1 below:

[4] Euclidean Distance is the measure between any two points in n-dimensional space as ordinarily measured by a ruler. It is given by the well-known Pythagorean formula for the distance between any two points P and Q. Assuming bi-dimensional space (given by F1(bark) and F2 (bark) values) the distance between points P and Q is given by the square root of the expression: $d(p,q) = (p1 - q1)^2 + (p2 - q2)^2$

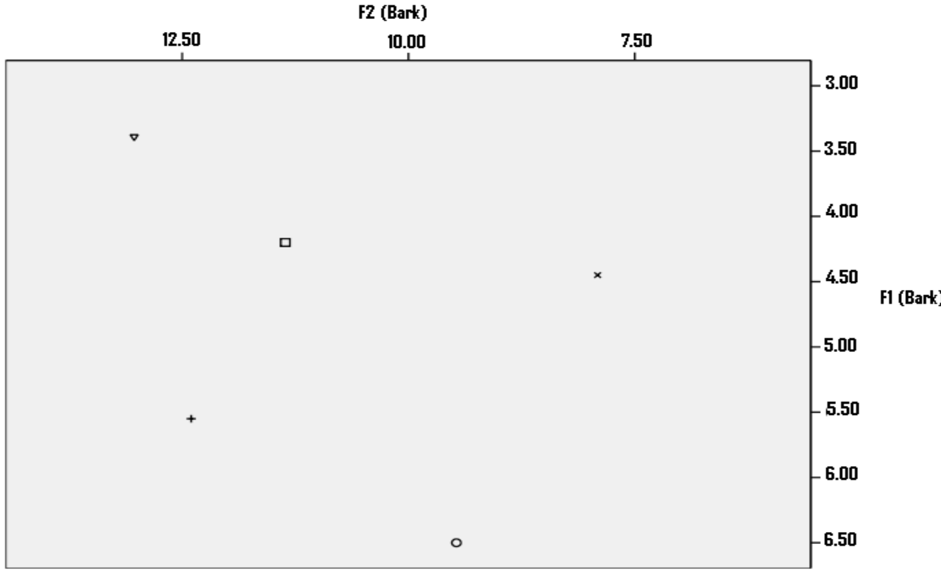


Figure 1: Vowel space of Akuntsú with average values for each vowel category over F1 (Bark) and F2 (Bark) dimensions.
Legend: ○ /a/, + /e/, x /o/, □ /ɨ/, and ▽/i/.

The Euclidean Distances between adjacent 'point vowels' are given in Table 3:

	/i-u/	/u-a/	/a-i/
D	*5.22*	*2.57*	*4.72*

Table 3: Bidimensional Euclidean distances between mean formant (bark) values for each one of the point vowels /a, i, u/.

By adding the distances in table 3 we find the value of 12.51, which can be used as an index of the overall dispersion of the language's vowel space. Comparing it with the same indexes computed over a 28-language sample reported in Livjin (2000) the value obtained for Akuntsú lies in the inferior part of the distribution for other languages with 5 vowels inventories (cf. figure 2 below). That is, for 5-vowels languages on which similar measures have been taken, the Akuntsú vowel space is a relatively crowded one. The value of 12.51 is not larger than that found for one language with a 4 vowels inventory in the Livjin sample, but it is larger than the values computed for languages with 10 and 8 vowels. Thus, the Akuntsú data corroborate the conclusions of Livjin (2000) study: there seems to be no definite trend for languages with larger vowel inventories to have larger vowel spaces on the F1/F2 (Bark) space. This is compatible and highly suggestive as Livjin (2000) acknowledges that the language may, in effect, have further mechanisms employed to support the discrimination of its vowel categories (say, vowel length). Below, figure 2 shows the approximate position of Akuntsú (colored circle) within a graphic showing the variation in the summed Euclidean distance *D* (ordinate) as a function of variation in inventory size (abscissa) for the Livjin (2000) sample.

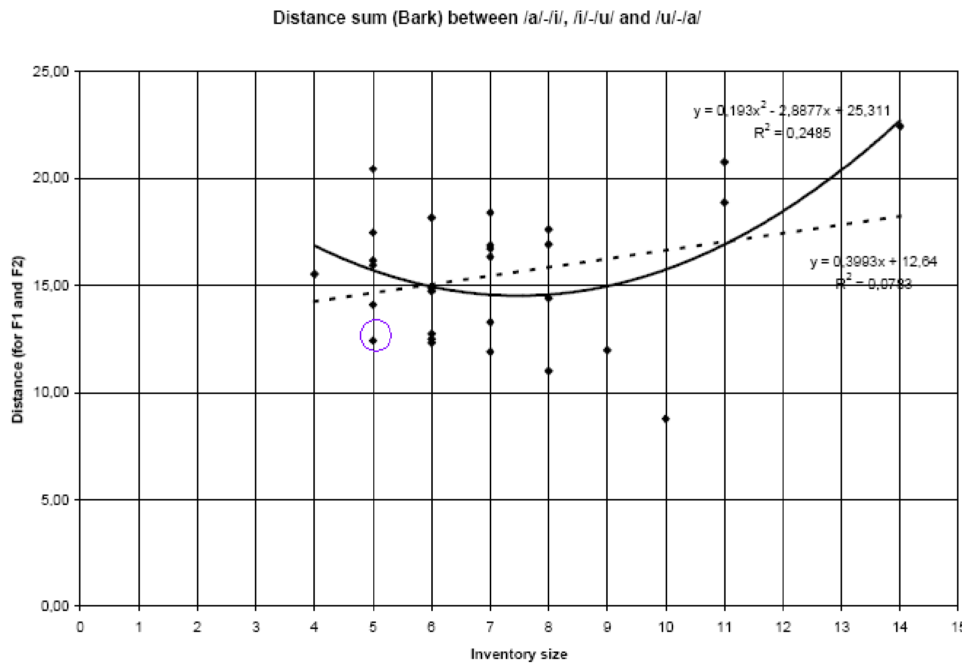


Figure 2: Plot of tendencies in the variation of vowel system acoustic dispersion (on F1 and F2 dimensions) as a function of vowel inventory size, with the approximate position of the Akuntsú system indicated by the colored circle. Taken from Livjin (2000).

3.2 PITCH AND PHONATION

A preliminary exploratory data analysis of the values on vowel pitch was carried, with and without speaker identity as a factor. The analysis was based on the pitch of vowels produced without audible creak.

	/i/	/e/	/a/	/ɨ/	/u/
Mean/Median	*151/149*	*156/149*	*112/106*	*174/168*	*156/154*

Table 4: Pooled central tendency values for vowel pitch

Mean/Median	/i/	/e/	/a/	/ɨ/	/u/
Male Consultant	*130/118*	*137/129*	*102/106*	*152/141*	*128/123*
Female Consultant	*194/194*	*193/193*	*132/132*	*216/216*	*213/213*

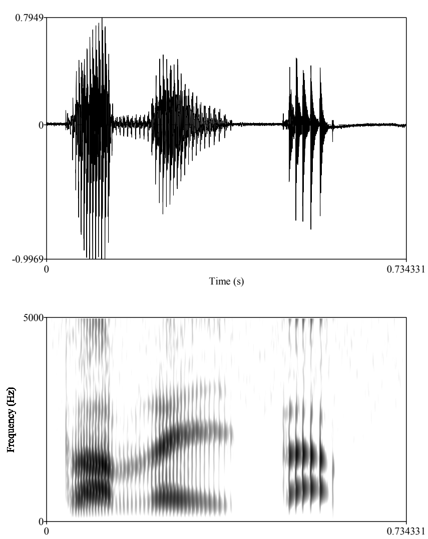
Table 5: Central tendency values for vowel pitch as a function of speaker identity/sex

As pointed out in section 3, vocalic segments in Akuntsú may be produced with non-modal creaky voice under certain conditions (cf. Figure 3 below). Laryngealization in Akuntsú is, then, produced with audible creak (cf. Blankenship 1997; for languages in which this is not the case). The presence of an adjacent glottal stop is one such conditioning factor, as tends to be the case cross-linguistically (Gordon 1998). However, as previous examples show (repeated here for the sake of clarity), the distribution of non-modal vowels is not restricted to those contexts in which a perturbation of modal voice is predictable from the presence of glottal stops, as follows:

(4)

[ḭˈpɛ̰bṵ]	‘its feather’	[ɔ̰ɾɔ̰ˈdḭta̰]	‘sloth (sp.)’
[ɔ̰ɾɔ̰ˈgʷḭ]	‘peanut’	[kɨˈbɛ̰g˺]	‘papaya’

In figure 3a and 3b below, we show respectively, the waveforms and spectrograms for two tokens of the lexical items /torotita/ [toroˈita] “rattlesnake” and /atap/ [aˈtap] “hair”, both produced with clear signs of laryngealization, especially in the last syllables, where an audible *creaky* quality is much stronger. A relative widening of the quasi-periodic pulses of glottal excitation, and a consequent drop in the average frequency of voicing (the main acoustic determinant of *pitch*) is evident from an inspection of these representations:



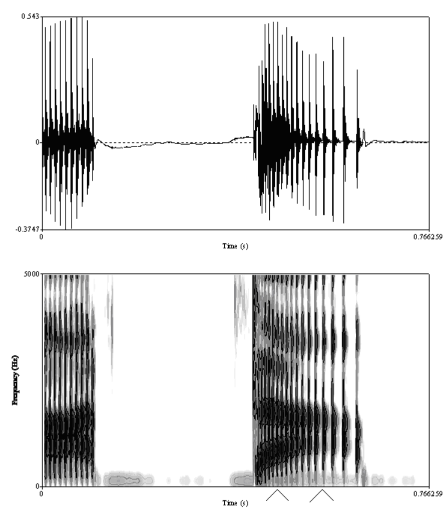


Figure 3a (on the left): Waveform and Spectrogram for the item [toroˈita] “rattlesnake”
Figure. 3b (on the right): Waveform and Spectrogram for the item [aˈtap] “hair”.

In order to probe into the phonetic nature of this form of pressed phonation, a few acoustic measures were employed. The first one, following a proposal by Kirk *et al.* (1984) tracks the period-to-period variation in cycle duration (*Jitter*). Laryngealized or pressed voice is taken to induce more jitter than modal voice. The durations of the cycles during a steady state vowel are manually taken, the means for consecutive periods are computed, pooled, and the Standard Deviation (*SD)* of the values is obtained. The larger the SD value, the larger the jitter or perturbation of the signal. Below we display token words with creaky voice and numeric values indicating the *SDs* obtained for consecutive syllables (the first four examples are from the male speaker, the remaining four from the female one):

	[ɨˈgḭ] 'water'	[iˈtḭɾḭ] 'two of them'	[kḭˈbɛ̰g˺] 'papaya'	[uˈttḁt˺] 'fire'
S.D.	0,107 – 0,526	0,09 – 4,176 – 0,166	0,049 – 4,799	0,077 – 3,875

Table 6: Standard deviations of the distributions of mean period length for voicing in the syllable nuclei of the indicated items; male consultant.

	[taˈpɛ̰] 'his way'	[a̰ˈta̰p] 'hair'	[ˈa̰gɛ̰ɾã̰̥] 'stand up'	[kɨˈpɛ̰] 'knife'
S.D.	0,117 – 9,228	0,263 – 4,738	0,92 – 2,535	0,197 – 2,454

Table7: Standard deviations of the distributions of mean period length for voicing in the syllable nuclei of the indicated items; female consultant.

A pattern in which larger *SDs* are obtained in the creaky-voiced syllable nuclei is evident from the above data. Another acoustic measure able to inform us on the nature of the phonatory processes involved in voice production is the H1-H2 coefficient[5] (Blankenship 1997; Klatt and Klatt 1990)**.** A high value of H1-H2 (i.e., a relative dominance of the first harmonic in the spectrum relative to the second one) indicates a long period of glottal abduction in each cycle of periodic voicing. Breathy voicing is then characterized by a large positive H1-H2 value, modal voicing has slightly slower values and pressed or laryngealized phonation is characterized by values closer to zero or negative values. A through and more systematic analysis of H1-H2 data wasn't carried out for this paper. Due to constraints as such pointed out in **3,** it was not possible to have a significant number of non-high vowels in controlled phonetic contexts required by such a study. Low vowels are used in to avoid the convergence of F1 and F0, especially for high-pitched female speech, which could make it harder to measure H1-H2 (cf. Blankenship 1997). Nevertheless, visual inspection of a few vowel spectra showing clearer harmonic structure was revealing. We took short-term spectra

[5] The H1-H2 coefficient is given by the difference in amplitude between the first and the second harmonics of a spectral section. It is taken to reflect the amount of time during a voicing period in which the glottal slit is open due to vocal fold medial abduction (cf. Blankenship 1997).

over 0.025 windows of voiced speech at the points indicated in the spectrogram at figure 3b. Figure 4 below shows two spectra, (a) the left on taken from the left point indicated in figure 3b; and (b) while the right spectrum was taken from the right point:

(a) **(b)**

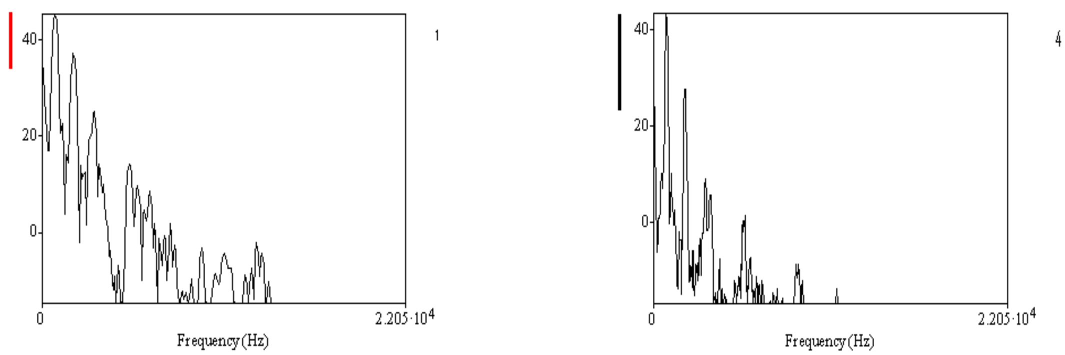


Figure 4: Two spectra taken at different moments in the production of the last vocoid in the word [a✂tap] at the points indicated in figure 3.b.

As the red bars at the upper left corner of the spectra show, in these two pulses of voicing excitation H1-H2 has a negative value, the amplitude of the second harmonic being higher than that of the first. This indicates that in a long part of each voicing period most of the glottal area is effectively reduced to zero due to medial adduction of the vocal folds. Another interesting fact is that, given the positions of the two spectra relative to what can be seen on the spectrogram and waveform at figure 3b, we have a scenario in which longer voicing periods (and then reduced F0) are associated to longer stretches of glottal adduction.

The set of data presented in this section indicates, then, that non-modal phonation in Akuntsú is produced via two distinct mechanisms: one that induces a larger period-to-period variation in the duration of the cycles of glottal excitation (creating a larger *jitter* in the speech signal) and one that enforces longer periods of vocal fold adduction or medial compression, though this latter mechanisms requires a more thorough investigation.

At this point it is worth pointing out that investigations such as these, probing into the details of the acoustic expression or phonetic implementation of underlying forms holds great promise for the study of a language's sound structure, as well as for a deeper understanding of sound change, even tough some linguists may fail to detect this relevance (cf. Ohala 1997).

As an example of the relevance of detailed phonetic studies to the more traditional concerns of linguists, we can cite John Kingston's work on tonogenesis in the Athabaskan family (cf. Kingston 2005). Through a consideration of the distinct, fine-grained consequences of glottalization on the phonatory activity of vowels, Kingston was able to elucidate the vexing question concerning the emergence of 'mirror' tone systems in Athabaskan languages, where, making a long story short, the presence of glottalized consonants induced low-level tones in some languages, but corresponding high-level tones in others.

Another property of laryngealization processes in Akuntsú which deserves further investigation is its apparent association with the phonetic implementation of word-level stress, as paradigms such as (4) show:

(5)

[kaˈɾo̰] 'necklace' [uˈka̰ɾo] 'my necklace'

The syllable where main stress falls seems to be invariably marked with pressed voice. When laryngealization occurs in any of the preceding syllables, it seems to be less strongly marked than in the stressed one.

4. CONCLUSIONS

The present paper has a number of distinct goals. In a direct improvement upon Aragon and Carvalho (2007) we provided a normalized version of the Akuntsú vowel space, as well as a perceptually grounded measure of global dispersion for this vowel space. In this work we also tried to present a deeper phonetic description of the phenomenon of creaky or laryngealized phonation in the Akuntsú language. Although there is no evidence for a lexical, underlying contrast between modally-voice and laryngealized vowels in Akuntsú, the strength with which pressed voice is realized in the language suggests a role for this distinguished phonetic event beyond that of variations in phonation setting due to the presence of glottal segments in the vicinity of the laryngealized vowels. There is evidence that in languages where non-modal phonation plays a role in the realization of relevant phonological oppositions, it is realized in a characteristically stronger way, when compared to languages in which it arises as an effect of allophonic perturbations on modal phonation (Blankenship 1997). In Akuntsú there is some evidence suggesting that it may play a role in stress contrasts.

REFERENCES

1. ARAGON, Carolina C. *Fonologia e aspectos morfológicos e sintáticos da língua Akuntsú*. MA. Thesis, University of Brasília, Brazil, 2008.
2. ARAGON, C. C. and Fernando O. Carvalho. Análise acústica das vogais orais da língua Akuntsú. In: *Revista da ABRALIN* v. 6, p. 41-55, 2007.
3. BLANKENSHIP, B. *The Time Course of Breathiness and Laryngealization in Vowels*. Doctoral Dissertation, University of California, Berkeley, 1997.
4. BOERSMA, P. & Weeninck, D. PRAAT. Doing Phonetics by Computer, 2006. www.praat.org
5. CABRAL, A.S.A.C. and C. Aragon. A posição da língua Akuntsú na família lingüística Tuparí. In: *Anais do IV Congresso Internacional da ABRALIN*, CD-Rom, pp. 1533-1539, 2005.
6. GORDON, M. The phonetics and Phonology of Non-Modal Vowels: a Cross-Linguistic Perspective. *Berkeley Linguistics Society* 24, 93-105, 1998.
7. KINGSTON, J. The Phonetics of Athabaskan Tonogenesis. In S. Hargus & K. Rice (Eds.), *Athabaskan prosody*. Amsterdam: John Benjamins Publishing, 137-184, 2005.
8. KIRK, P., P. Ladefoged and J. Ladefoged. Using a Spectrograph for Measures of Phonation Type in Natural Languages. *UCLA Working Papers in Phonetics* 59, 102-113, 1984.
9. KLATT, D. and L. Klatt. Analysis, Synthesis and Perception of Voice QualityVariations Among Male and Female Talkers. *Journal of the Acoustical Society of America 87,* 820-857, 1990
10. LIVIJN, P. Acoustic Distribution of Vowels in Differently Sized Inventories – Hot Spots or Adaptive Dispersion?. *XIIIth Swedish Phonetics Conference Proceedings*, p. 93-96. Skövde, Sweden, 2000.
11. OHALA, J. Phonetics in Phonology. *Proceedings of the 4th Seoul International Conference on Linguistics (SICOL)*, p. 45-50, 1997.
12. TRAUNMÜLLER, H. Analytical expressions for the tonotopic sensory scale. *Journal of the Acoustical Society of America* 88, 97-100, 1990.

RESUMO: Este artigo apresenta uma análise preliminar das vogais laringalizadas na língua Akuntsú, falada por seis índios monolíngues no sudeste do estado de Rondônia, Brasil. Os espaços vocálicos são apresentados em termos de uma escala psico-acústica padronizada, incluindo outras informações quantitativas. Uma medida global de dispersão do sistema vocálico é apresentada, e o valor encontrado para o Akuntsú é comparado com o de outras línguas de uma amostra independente. Resultados sugerem que a fonação não-modal nesta língua pode ser caracterizada por pelo menos dois mecanismos articulatórios distintos e seus efeitos acústicos.
PALAVRAS-CHAVE: Língua Akuntsú; fonação não-modal; Espaço Vocálico; Acústica

ABSTRACT: This paper aims to present a preliminary analysis of creaky or laryngealized vowels in Akuntsú spoken by only six monolingual indians on the Omerê River, at the Brazilian western state of Rondônia, Brazil. Vowel spaces are re-expressed in terms of a standard psychoacoustic scale and further quantitative information is addressed. A global measure of acoustic dispersion is presented and the value obtained for the Akuntsú vowel space is compared to similar values obtained from an independent sample of languages. Results suggest that non-modal phonation in Akuntsú can be characterized as the effect of at least two distinct articulatory mechanisms and their differing acoustic effects.
KEYWORDS: Akuntsú language; Non-modal phonation; Vowel space; Acoustics.