International Journal of Primatology, Vol. 24, No. 1, February 2003 (© 2003)

Vocal Production by a Language-Competent *Pan paniscus*

Jared P. Taglialatela,[1,2] Sue Savage-Rumbaugh,[1] and Lauren A. Baker[1]

Received September 18, 2001; revised Janaury 29, 2002; accepted March 11, 2002

Human spoken language and nonhuman primate vocalization systems have traditionally been regarded as qualitatively different from one another with respect to their semanticity and the way in which individuals acquire and utilize these signals. However, recent studies of the vocal behaviors of both captive and free-ranging monkeys and apes suggest that this dichotomy may not be unequivocal. We examined the vocalizations produced by a linguistically-competent adult male bonobo (Pan paniscus) named Kanzi. We analyzed his vocalizations during communicative interactions with humans in order to determine whether they vary systematically according to the semantic context in which they are produced. We determined semantic contexts based upon a vocalization's co-occurrence with predefined behavioral correlates. Spectrographic and statistical analyses revealed that acoustic structure is similar among the vocalizations that occurred within a specific semantic context and structural differences are evident between the vocalizations produced in different contexts. The results provide evidence that, during communicative interactions with humans, Kanzi modulates his vocal output on both the temporal and spectral levels.

KEY WORDS: primate vocalization; language; speech; bonobo; Kanzi; *Pan paniscus*.

---

[1] Language Research Center, Georgia State University, 3401 Panthersville Road, Decatur, Georgia 30034.

[2] To whom correspondence should be addressed;e-mail: jared@gsu.edu.

INTRODUCTION

Traditionally human spoken language and nonhuman primate vocalization systems have been considered fundamentally different from one another. Whereas human spoken languages are learned, under voluntary control, referential, and grammatically organized, nonhuman primate vocalization systems have been described as genetically preprogrammed, involuntary, and bound to the motivational state of the signaler (Lieberman, 1998). However, studies of the vocal behaviors of both captive and free-ranging monkeys and apes suggest that this dichotomy has been overstated. Many nonhuman primate species behave as though the vocalizations of their conspecifics refer to objects, individuals, or events in their environments (Hauser, 1998; Seyfarth *et al.*, 1980; Struhsaker, 1967; Zuberbuhler, 2000; Zuberbuhler *et al.*, 1999). For example, Zuberbühler *et al.* (1999) conducted a series of playback experiments in order to determine if diana monkeys (*Cercopithecus diana diana*) attend to the meanings of alarm vocalizations produced by conspecifics or simply to the acoustic characteristics of the calls. They presented groups of monkeys with 2 calls in succession, separated by a 5-min period of silence. The playback stimuli included four types of vocalizations: male diana monkey alarm calls in response to leopards, male diana monkey alarm calls in response to eagles, and the calls produced by the two predators. They found that the monkeys readily transferred habituation across acoustic, but not semantic features. In other words, when monkeys were presented with conspecific alarm vocalizations for eagles followed by the shrieks produced by crowned eagles, the monkeys did not dishabituate. However, when the shrieks produced by crowned eagles were preceded by monkey alarm calls for leopards the monkeys did not transfer their habituation. The results suggest that diana monkeys attend to the underlying meaning of the vocalizations.

In addition, recent data indicate that many nonhuman primate species learn to produce certain calls in specific contexts and that learning may also play a role in shaping the structure of some of the vocalizations they produce (Cheney and Seyfarth, 1990; Snowdon, 1994; Fisher *et al.*, 1998; Hopkins and Savage-Rumbaugh, 1991; Marshall *et al.*, 1999; Masataka, 1988; McCowan and Newman, 2000; Mitani *et al.*, 1992; Maeda and Masataka, 1987). For example, Marshall *et al.* (1999) recently compared the pant-hoot vocalizations produced by two captive groups of male chimpanzees (*Pan troglodytes*). They reported that the structures of the calls produced by the members of each group have converged as a result of vocal learning.

Furthermore, data from studies that have employed conditioning paradigms to determine whether nonhuman primates can be trained to produce or sequester vocalizations indicate that primate vocal production may also contain a volitional component (Aitken, 1981; Aitken and Wilson, 1979; Pierce, 1985; Randolph and Brooks, 1967; Sutton *et al.*, 1973, 1978, 1981). Although the extent of volitional control remains unclear, Jürgens (1988) suggested that any limitation may be attributable to the fact that, unlike humans, monkeys lack direct connections between the primary motor cortex and the laryngeal motorneurons . However, recent neurological evidence suggests that the neocortical area that is homologous to Broca's area in humans participates in the generation and control of Japanese monkey vocalizations (Gemba *et al.*, 1995).

Considered collectively, these data blur the traditional boundary that separates human speech from all other primate vocal communicative systems. Further evidence of this continuity is provided by data from studies to discern whether apes possess the potential for linguistic competency. The results of these investigations indicate that chimpanzees and bonobos are able to acquire and to utilize an artificially constructed lexical communication system, and to comprehend novel spoken English requests when reared in a language-enriched environment (Brakke and Savage-Rumbaugh, 1995a,b; Savage-Rumbaugh, 1986; Savage-Rumbaugh *et al.*, 1993; Sevcik and Savage-Rumbaugh, 1994). The question then naturally arises, if apes are capable of spontaneously acquiring linguistic competency in a nonvocal domain, might they also have the potential for similar capacities in the vocal domain?

Some reserchers have described the production and use of vocalizations by chimpanzees and bonobos in captivity and in the wild (Bermejo and Omedes, 1999; De Waal, 1988; Goodall, 1986; Hallberg *et al.*, 1999; Hauser and Wrangham, 1987; Hohman and Fruth, 1994; Hopkins and Savage-Rumbaugh, 1991; Marler, 1969; Marshall *et al.*, 1999; Mitani *et al.*, 1992; Mori, 1982, 1983; van Hoof, 1972). DeWaal (1988) examined the communicative repertoire of 10 captive bonobos in three social groups at the San Diego Zoological Garden and classified their vocal emissions into 12 broad vocalization categories. Of the 12, the most common vocalizations were food peeps, which, "are used to draw attention to and 'comment' on objects, food, and events in the environment" (De Waal, 1988). High structural variability existed among the calls in the group and among the consecutive vocalizations of a single caller. DeWaal (1988) suggested that the structural variations may in fact be meaningful to conspecifics.

Hohman and Fruth (1994) noted similar structural variability in the high-hoots (De Waal, 1998) produced by a group of wild bonobos in the Lomako forest. They examined the structure and use of the distance calls and suggested that they, "may be the major device to regulate and to maintain the social network of the community" (Hohman and Fruth, 1994). Individuals in close spatial proximity of one another alternate their call production and adjust the structural characteristics of their emissions to correspond with the vocalizations of conspecifics. This finding is consistent with De Waal's (1988) report that bonobos similarly adjust their calls when producing contest hoots and screams.

Further Hopkins and Savage-Rumbaugh (1991) reported instances of vocal learning by Kanzi at 5 years of age. They compared Kanzi's vocal repertoire with that of a socially housed group of bonobos at the Yerkes field station. Their results indicate that in addition to the species-typical vocalizations produced by both the members of the socially housed Yerkes bonobos and by Kanzi, four additional vocalization types were produced exclusively by Kanzi. These vocalizations occured in the context of human questions, comments, and gestures. Hopkins and Savage-Rumbaugh (1991) concluded that the structure and use of Kanzi's vocalizations were altered due to his unique, language-enriched rearing experience.

We sought to evaluate systematically the vocal utterances produced by Kanzi, now an adult, during everyday communicative interactions with human caregivers and researchers. The exchanges share characteristics with human conversation as Kanzi participates in turn-taking and responds to, and produces requests (Benson *et al.*, 2002). Many primatologists have recognized the importance of studying vocal exchanges between conspecifics and have noted their similarity to human conversation (Biben *et al.*, 1986; Biben and Symmes, 1991; Masataka and Biben, 1987; Smith *et al.*, 1982; Snowdon and Cleveland, 1984; Sugiura, 1993, 1998, 1998; Symmes and Biben, 1988). During exchanges with humans, Kanzi's vocalizations are often paired with spoken words produced by caregivers and researchers, as well as with his own gestures to lexigrams, photographs, and objects. These concomitant behaviors make it possible to objectively define the semantic context in which Kanzi produces a vocalization. We hypothesized that Kanzi regulates his vocal output by producing structurally distinct sounds that are temporally associated with distinct semantic contexts. To directly evaluate the hypothesis, we examined the structure of the vocalizations Kanzi produced during interspecific linguistic interactions with humans.

METHOD

Vocalization Collection Procedure

We categorized Kanzi's vocalizations according to the semantic context in which he produced them. We used a video archive as the source of behavioral data, and considered *ca.* 100 h of video footage for the analysis. We scanned video data for pertinent segments in which Kanzi was interacting with one or more humans. These scenarios include experimental test sessions and informal communicative interactions typical of Kanzi's daily experience. Kanzi frequently produces vocalizations during these interactions, and his vocalizations are temporally associated and interspersed with speech by humans. Figure 1 is an example of one such exchange. It illustrates the spoken English utterances of a researcher, and the vocalizations produced concurrently by Kanzi during a single communicative interaction. We collected Kanzi's vocalizations during similar interactions and classified them according to the semantic context in which they were produced. Semantic context is operationally defined according to predetermined behavioral correlates, and we assigned a vocalization to a semantic context based on its co-occurrence with one or more of these behavioral correlates (Table I). The behavioral correlates included lexigram use, gesture to objects, and response to polar interrogatives: yes – no questions. For example, if Kanzi vocalized while gesturing to the lexigram symbol for a particular object, we labeled the vocalization with the name of that object. Similarly, if Kanzi vocalized following a human speaker who had indicated the name of a particular object,we labeled the vocalization with the name of that object. We considered only vocalizations produced in conjunction with one or more behavioral correlates and characterized by only a single semantic context.

The rationale for this procedure is twofold. First, we assigned vocalizations to a semantic context according to their association with $\geq$1 of the predetermined behavioral correlates in order to safeguard against potential observer biases based on the perceptual properties of the sound. In addition, a subset of the vocalizations ($N = 19$) analyzed in this study were scored by a second independent observer who viewed edited video segments in which the vocalizations were omitted and replaced with a tone. This was done to further ensure that Kanzi's behavior defined a vocalization's semantic context rather than the observer's perceptual impression of the call. The second observer's assignments of vocalizations were completely consistent with the first observer's classifications.

Second, we included only vocalizations that could be assigned to a single semantic context. For example, if Kanzi vocalized and gestured to the lexigram for one item while eating a second item, that vocalization could not be unambiguously assigned to a single semantic context and was excluded from the analysis. This was done in order to ensure that vocalizations were consistently classified according to their semantic context.

Vocalization Analysis

Once the collection procedure was complete, we quantified each vocalization in order to compare the acoustic properties across and within semantic categories. We digitized vocalizations and analyzed them with a

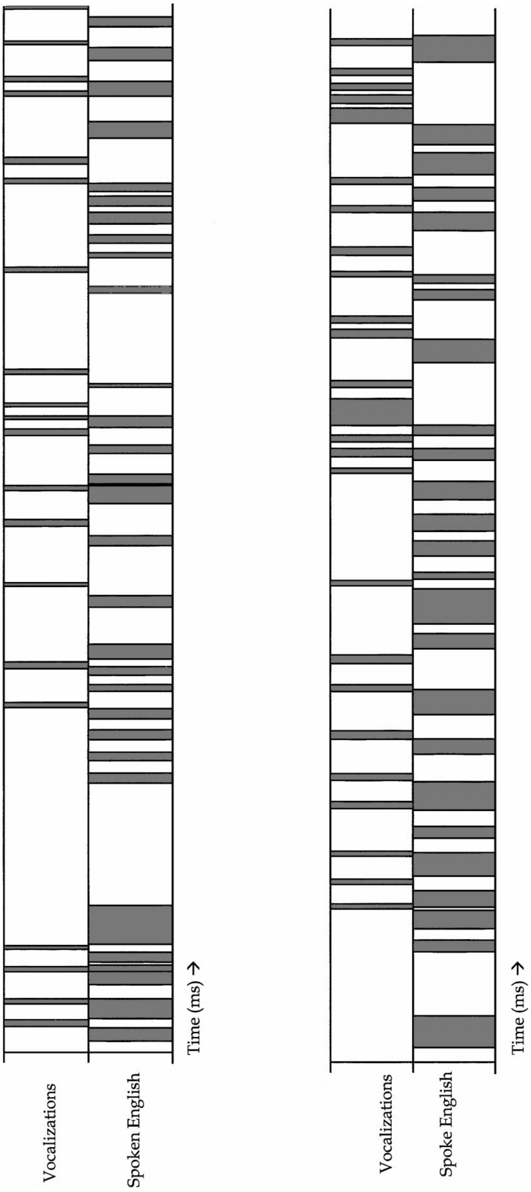

Fig. 1. The two traces display the initial time and duration for vocal utterances produced by Kanzi and spoken English produced by a researcher during a 2-min communicative exchange. Note the alternating pattern and the association of speech and vocal production.

Table I. Vocalization collection criteria

Behavioral correlates	Example
Bonobo vocalizes and Gestures to lexigram Gestures to object Gestures to photograph	Vocalization in conjunction with gesture to lexigram for "grapes" Semantic Context: grapes
Bonobo vocalizes following a human speaker who is talking to the ape or another individual	Vocalization after human speaker said, "Kanzi would like another banana." Semantic Context: banana
Bonobo vocalizes and directly attends to a food item or object	Vocalization while Kanzi is presented with juice. Semantic Context: juice
Bonobo vocalizes in response to direct polar query posed by a human participant	Vocalization following human speaker asking, "Kanzi, are you ready to listen?" Semantic Context: yes

Power Macintosh G3 computer (sample rate = 44.1 kHz, 16 bit precision) and a signal analysis software program (Canary v1.2.4, The Bioacoustics Research Program, Ithaca, NY). We constructed sound spectrograms of each vocalization (FFT size = 1024 points) and made 4 on-screen measurements (Fig. 2) and performed 12 calculations yielding a total of 16 acoustic variables for each vocalization (Table II).

Data Screening and Statistical Analysis

We removed univariate and multivariate outliers to reduce the probability of committing Type I and/or Type II errors due to extreme values.

Table II. Acoustic variables

Acoustic variable	Description
Onset	Initial frequency of vocalization (kHz)
Peak	Highest frequency of vocalization (kHz)
Endpoint	Endpoint frequency of vocalization (kHz)
Minimum	Lowest frequency of vocalization (kHz)
Duration	Endpoint time minus onset time (ms)
Onset−peak	Peak time minus onset time (ms)
Peak−endpoint	Endpoint time minus peak time (ms)
Frequency mean	(Onset + peak + endpoint)/3 (kHz)
Peak − minimum range	Maximum frequency minus minimum frequency (kHz)
Onset − endpoint range	Onset frequency minus endpoint frequency (kHz)
Onset − peak range	Maximum frequency minus onset frequency (kHz)
Endpoint − peak range	Maximum frequency minus endpoint frequency (kHz)
Peak location	Peak time (ms) - onset time (ms)/duration (ms)
Minimum location	Minimum time (ms) - onset time (ms)/duration (ms)
Slope, onset − peak	Peak (kHz) - onset (kHz)/peak time (ms) - onset time (ms)
Slope, peak-endpoint	Peak (kHz) - endpoint (kHz)/peak time (ms) - endpoint time (ms)

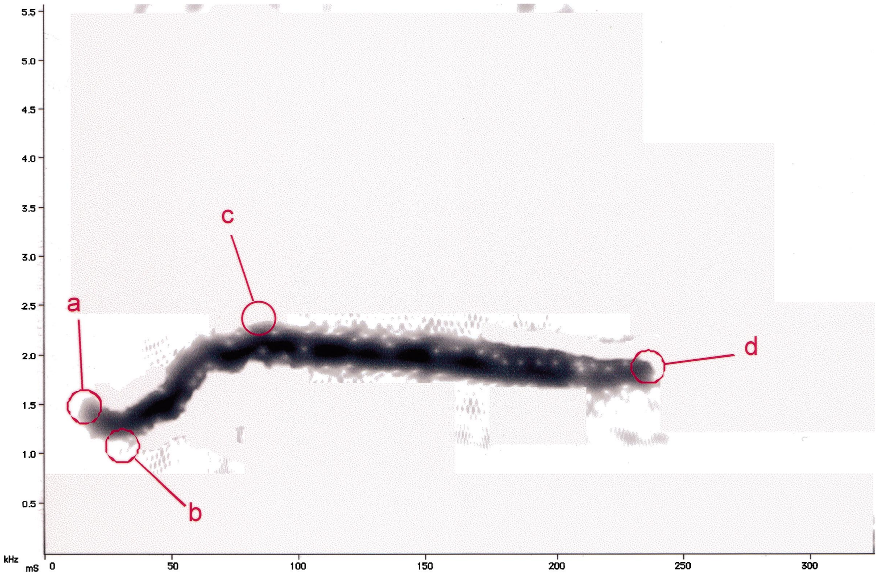

Fig. 2. Sample Spectrogram. On-screen measurement of acoustic variables, onset (a), minimum (b), peak (c), end (d). For each of the 4 measured acoustic variables, we recorded a frequency (kHz) and time (ms) value and used them to calculate the 12 additional acoustic variables (Table II) and (Table III).

Following the recommendation of Tabachnick and Fidell (1996), we identified standardized scores over 3.29 as possible univariate outliers and removed them from the data set. In addition, cases that have an unusual pattern of scores are potential multivariate outliers. For example, a vocalization from a given semantic context may have combinations of acoustic measurements that are vastly different from the other vocalizations in that group. Therefore, we obtained Mahalanobis distance (— distance of each case from the centroid of the group—) for each vocalization via linear regression with a dummy DV (case number) for each vocalization group and removed any case identified as a potential multivariate outlier from the data set.

We conducted a univariate analysis of variance to identify differences among semantic contexts for each acoustic variable. We performed Tukey's Honestly Significant Difference (HSD) tests to identify differences ($p < .05$) between pairs of semantic contexts for each acoustic variable.

We employed a multinomial logistic regression (MLR) in order to determine whether the semantic context in which vocalizations were produced could be predicted based upon their values for the acoustic variables. Multinomial logistic regression answers the same questions as discriminant function analysis (DFA); however, MLR is more flexible and has no assumption about the distribution of predictor variables and the predictors need not be normally distributed (Tabachnick and Fidell, 1996). Multinomial logistic regression, like DFA, provides classification of members of a dependent variable based on weighted contributions of predictor variables. Classification indexes, model goodness of fit, and relative contributions of each predictor variable to the overall fit of the model are calculated (Dattalo, 1994; Derose, 1991; Tabachnick and Fidell, 1996). The outcomes of DFA and MLR for the same data set are not significantly different (Derose, 1991; Meshbane, 1996). Therefore, we performed a direct multinomial logistic regression via SPSS 10.0 to assess whether the semantic contexts in which the vocalizations were produced could be predicted on the basis of the relevant acoustic variables.

RESULTS

Descriptives and Univariate Analysis

For purposes of statistical reliability, we analyzed only semantic contexts with 10 or more vocalizations. We named the four vocalization groups that fulfilled this criterion according to the semantic context that characterized the members of that group: banana ($N = 10$), grape ($N = 11$), juice ($N = 14$), and yes ($N = 25$). Data screening identified three cases in the juice group as outliers, and we excluded them from further analysis. The classification procedure was extremely conservative to ensure that vocalizations were grouped based on the behavioral correlates. However, it is possible that a few vocalizations may have been misclassified. This would result in very different values for acoustic variables for that exemplar, resulting in the exclusion of that call based on the statistical rationale indicated. Nevertheless, we contend that the observed variation in nonhuman primate calls is very important. In fact, all of Kanzi's calls included in this study can be described as members of a single previously defined vocal class, eg., De Waal's peeps (1988). De Waal (1988) himself noted the great variation in these calls, and this variation is what we examined.

Figure 3 depicts, for each vocalization group, the mean values for the measured acoustic variables onset, peak, endpoint, and minimum and their relative locations in the vocalizations, e.g., peak and its relative location with respect to onset, endpoint, and minimum.

After Bonferoni's adjustment ($.05/16 = .003$) there are significant differences among semantic context for 6 of the 16 acoustic variables (Table III). In addition, semantic context accounts for a considerable proportion of variance for several acoustic variables. We performed Tukey's HSD to identify which semantic contexts differed for each acoustic variable. Several semantic context pairs differ on >1 acoustic variable (Table IV).

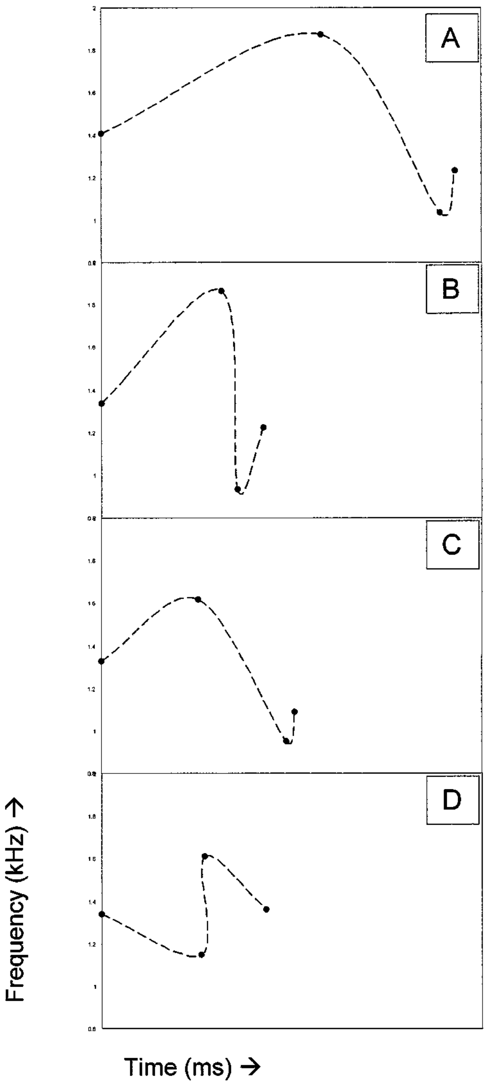

Fig. 3. Composite representations of each of the 4 vocalization categories. Data points represent the mean values for the acoustic variables—onset, peak, endpoint, and minimum—for each of the 4 vocalization groups: banana (A), grape (B), juice (C), and yes (D).

Table III. Univariate statistics[a]

Acoustic variable	Banana		Grape		Juice		Yes		F	p	Eta Squared
Onset (SD)	1.41	(.20)	1.34	(.16)	1.29	(.33)	1.34	(.22)	.44	.728	.024
Peak (SD)	1.87	(.31)	1.87	(.36)	1.60	(.31)	1.61	(.26)	3.52	.021	.166
Endpoint (SD)	1.23	(.26)	1.22	(.44)	1.04	(.31)	1.36	(.29)	2.50	.070	.124
Minimum (SD)	1.04	(.21)	.94	(.23)	.91	(.29)	1.15	(.23)	3.54	.021	.167
Duration (SD)	372.06	(144.33)	170.10	(32.86)	216.22	(51.19)	172.98	(50.90)	19.29	.000	.522
Onset – peak (SD)	146.69	(106.46)	74.66	(25.50)	81.03	(47.52)	81.83	(39.31)	3.90	.014	.181
Peak – endpoint (SD)	225.36	(100.95)	95.43	(36.86)	135.19	(75.16)	91.15	(39.65)	12.40	.000	.412
Frequency mean (SD)	1.50	(.21)	1.48	(.28)	1.31	(.28)	1.44	(.24)	1.27	.296	.067
Peak – minimum range (SD)	.84	(.33)	.93	(.24)	.69	(.26)	.46	(.13)	14.23	.000	.446
Onset – endpoint range (SD)	.23	(.23)	.40	(.15)	.27	(.28)	.17	(.12)	3.97	.013	.183
Onset – peak range (SD)	.47	(.20)	.53	(.26)	.31	(.17)	.27	(.15)	6.30	.001	.263
Endpoint – peak range (SD)	.64	(.32)	.64	(.31)	.55	(.28)	.25	(.13)	10.70	.000	.377
Peak location (SD)	.38	(.16)	.46	(.21)	.40	(.23)	.47	(.18)	.69	.687	.037
Minimum location (SD)	.79	(.27)	.53	(.47)	.88	(.28)	.45	(.44)	3.81	.015	.177
Slope, onset – peak (SD)	.004	(.003)	.008	(.004)	.006	(.006)	.005	(.006)	0.92	.437	.050
Slope, peak – Endpoint (SD)	–.003	(.002)	–.008	(.006)	–.005	(.002)	–.003	(.002)	9.04	.000	.339

[a] Univariate statistics (mean, SD), ANOVA, and Eta Squared values for the acoustic variables measured for each of the 4 vocalization groups: banana, grape, juice, and yes ($df = 3, 56$). This method is similar to those previously employed by investigators to compare vocalizations quantitatively and to evaluate hypothesized structural differences based on caller identity and/or the context in which the vocalizations were produced (Norcross and Newman, 1997; Norcross *et al.*, 1999; Gouzoules and Gouzoules, 2000)

Table IV. Post hoc tests

Acoustic variable	banana–grape	banana–juice	banana–yes	grape–juice	grape–yes	juice–yes
Onset						
Peak						
Endpoint						*
Minimum						*
Duration	*	*	*			
Onset-peak	*	*	*			
Peak–endpoint	*	*	*			
Frequency mean						
Peak–minimum range			*		*	*
Onset–endpoint range					*	
Onset–peak range			*		*	
Endpoint–peak range			*		*	*
Peak location						
Minimum location						*
Slope, onset-peak						
Slope, peak–endpoint	*			*	*	

Multinomial Logistic Regression

We performed a direct multinomial logistic regression analysis via SPSS 10.0 to assess prediction of membership in one of 4 semantic contexts on the basis of the 6 acoustic variables with significant differences among groups. A comparison of the full model with this set of acoustic variables against a constant-only model is significantly reliable, $\chi^2(18, 57) = 80.80$, $p < .0001$, indicating that the predictors, as a set, reliably distinguish the 4 semantic contexts.

The overall semantic context prediction rate is 74%. Prediction rates for banana (80%), grape (82%), and yes (84%) are robust; however, the prediction rate for juice is marginal (36%). The acoustic predictors included in the model did not reliably distinguish vocalizations in the juice category from vocalizations in other categories. In order to find the set of predictor variables that could best predict group membership for all 4 semantic contexts, we added acoustic variables with $\mathrm{p} < .05$, as indicated in the initial analysis of variance, to the original model. Two of the acoustic variables—(peak-minimum range and peak-endpoint)—failed the tolerance test for multicollinearity and were not included in the revised model.

We performed a direct multinomial logistic regression analysis to assess prediction of group membership in one of the 4 semantic contexts based on the 9 acoustic variables: peak, minimum, duration, onset-peak, onset-endpoint range, onset-peak range, endpoint-peak range, minimum location, slope; peak-endpoint. A test of the revised model, including all

9 acoustic variables, against a constant-only model is statistically reliable, $\chi^2(27, 57) = 100.79$, $p < .0001$, indicating that this set of predictors achieves reliable discrimination among the 4 semantic categories. The variance in group accounted for is satisfactory (McFadden's Rho = .68). Overall prediction rate is 88%. Prediction rates for each group are robust with banana = 90%, grape = 91%, juice = 82%, and yes = 88%.

DISCUSSION

Visual inspection of Fig. 3 reveals that the mean values for the measured acoustic variables for each vocalization group are structurally distinct from one another. This is corroborated by the results of the univariate statistical analysis, which indicates that both temporal properties, e.g. the duration of the vocalizations, and spectral characteristics, e.g. peak frequencies, account for a considerable amount of the variance between vocal categories (Table III), and by the results of the post-hoc tests (Table IV), which indicate that for each pair of vocalization groups there are differences between them for ≥ 1 acoustic variables. For example, trace A in Fig. 3 appears to be nearly twice as long as the other 3. In addition, trace D appears to have a different shape from the others because the minimum frequency preceded the peak frequency in the exemplars for this vocal category. The results of the multinomial logistic regression indicate that it is possible to predict the semantic context in which Kanzi produced a vocalization based on its acoustic structure. Furthermore, our results suggest that the observed structural differences among vocalizations produced in different semantic contexts cannot be attributed to only one or two acoustic variables. Instead, the vocalizations vary as a result of a suite of acoustic variables, with certain variables distinguishing any one group from another, and other variables distinguishing a third group from the fourth (Table IV). The data indicate that the observed structural variability is predictable, and not merely the result of random variation in the spectral or temporal characteristics of Kanzi's vocal utterances.

Given that we included vocalizations from only 4 semantic contexts in the analysis, a skeptic might claim that the sounds are simply the result of differences in Kanzi's emotional state. If this were the case, then Kanzi's emotional state, at the very least, enables him to produce distinct sounds in distinct semantic contexts. However, details of the collection methods and results render this explanation unlikely. First, the conditions in which all of the vocalizations were produced varied greatly with respect to how recently Kanzi had eaten, gone outside, or what his preferred food item was at any given time. For example, a vocalization produced by Kanzi and classified as the semantic context banana might have been produced while he was eating bananas, gesturing to the lexigram for bananas, or in response to a human telling him there were no bananas (Table I). Moreover, if the idea of banana itself simply elicited an emotional response in Kanzi that resulted in a vocalization, we might expect that Kanzi would vocalize every time he saw a banana, touched the lexigram for banana, or heard the spoken word, banana. However, this is not the case. For example, Kanzi is often asked to select a photograph or lexigram that corresponds to a spoken word. He is able to complete this task, and often does so while remaining completely silent.

Secondly, vocalizations in the semantic context yes were produced by Kanzi while his affective state varied greatly because the vocalizations in this group were produced while he was eating, moving, playing, working, etc., and were not consistently associated with a single food item. Accordingly, if Kanzi's emotional state were solely responsible for the observed structural differences between vocalizations, one would expect to observe high structural variability for the exemplars in this semantic context. However, as reported in Table II., the standard deviations for the acoustic variables in the category yes were small. Although Kanzi's emotional system most likely plays a role in his vocal behavior, this factor alone cannot account for our results.

The data indicate that Kanzi produces distinct vocalizations that vary systematically according to the semantic context in which he produces them. These results coincide with data concerning the vocal competencies and communicative behaviors of bonobos in the wild and in captivity (De Waal, 1988; Goodall, 1986; Hauser and Wrangham, 1987; Hohman and Fruth, 1994; Hopkins and Savage-Rumbaugh, 1991; Marshall *et al.*, 1999; Mitani *et al.*, 1992; Mori, 1982, 1983). However, our results extend these findings significantly for they provide evidence that a language-competent bonobo modulates his vocal output, on both temporal and spectral levels, during interspecific linguistic interactions with humans. These conversations are similar to the close-range, affiliative vocal exchanges that have been studied in both New and Old World monkey species (Biben *et al.*, 1986; Biben and Symmes, 1991; Masataka and Biben, 1987; Smith *et al.*, 1982; Snowdon and Cleveland, 1984; Sugiura, 1993, 1998; Symmes and Biben, 1988). Newman, *et al.* (1983) suggested that the variation of structural elements in the affiliative vocalizations of nonhuman primates may function to convey meaning to conspecifics. Given Kanzi's well-documented linguistic competencies, (Savage-Rumbaugh *et al.*, 1993), and our findings, we suggest that the structural variability in Kanzi's vocal utterances may serve a similar function. Our future work will focus on evaluating this potential for semantic vocal communication.

ACKNOWLEDGMENTS

We thank John D. Newman for providing procedural guidance with vocalization quantification, Roger Bakeman for consultation regarding statistical analyses, and Duane M. Rumbaugh and the faculty and staff at the Language Research Center for their assistance with this project. We also thank John D. Newman, Herbert L. Roitblat, and two anonymous reviewers for comments on an earlier draft of the manuscript. This work was supported by an NIH grant to S. Savage-Rumbaugh. (NICHD-06016).

REFERENCES

Aitken, P. G. (1981). Cortical control of conditioned and spontaneous vocal behaviour in Rhesus monkeys. *Brain Lang.* 13: 171–184.

Aitken, P. G., and Wilson, D. A. (1979). Discriminative vocal conditioning in Rhesus monkeys: Evidence for volitional control? *Brain Lang.* 8: 227–240.

Benson, J., Fries, P., Greaves, W., Iwamoto, K., Savage-Rumbaugh, S., and Taglialatela, J. (2002). Confrontation and support in bonobo-human discourse. *Functions Lang.* 9: 1–33.

Bermejo, M., and Omedes, A. (1999). Preliminary vocal repertoire and vocal communication of wild bonobos (*Pan paniscus*) at Lilungu (Democratic Republic of Congo). *Folia Primatol.* 70: 328–357.

Biben, M., and Symmes, D. (1991). Playback studies of affiliative vocalizing in captive squirrel monkeys: Familiarity as a cue to response. *Behaviour* 117(1/2): 1–19.

Biben, M., Symmes, D., and Masataka, N. (1986). Temporal and structural analysis of affiliative vocal exchanges in squirrel monkeys (*Saimiri sciureus). Behaviour* 98: 259–273.

Brakke, K. E., and Savage-Rumbaugh, E. S. (1995a). The development of language skills in bonobo and chimpanzee: i. Comprehension. *Lang. Commun.* 15: 121–148.

Brakke, K. E., and Savage-Rumbaugh, E. S. (1995b). The development of language skills in *Pan*: ii. Production. *Lang. Commun.* 16: 361–380.

Cheney, D. L., and Seyfarth, R. M. (1990). *How Monkeys See the World: Inside the Mind of Another Species*, University of Chicago Press, Chicago.

Dattalo, P. (1994). A comparison of discriminant analysis and logistic regression. *J. Soc. Sci. Res.* 19(3/4): 121–144.

De Waal, F. B. M. (1988). The communicative repertoire of captive bonobos, *Pan paniscus,* compared to that of chimpanzees. *Behaviour* 106: 183–251.

Derose, D. (1991), Comparing Classification Models from Multinomial Logistic Regression and Multiple Discriminant Analysis. New York University Press, New York.

Elowson, A. M., and Snowdon, C. T. (1994). Pygmy marmosets, *Cebuella pygmaea,* modify vocal structure in response to changed social environment. *Anim. Behav.* 47: 1267–1277.

Fisher, J., Hammerschmidt, K., and Todt, D. (1998). Local variation in Barbary macaque shrill barks. *Anim. Behav.* 56: 623–629.

Gemba, H., Miki, N., and Sasaki, K. (1995). Cortical field potentials preceding vocalization and influences of cerebellar hemispherectomy upon them in monkeys. *Brain Res.* 697: 143–151.

Goodall, J. (1986). *The Chimpanzees of Gombe: Patterns of Behavior*, Harvard University Press, Cambridge, MA.

Gouzoules, H., and Gouzoules, S. (2000). Agonistic screams differ among four species of macaques: The significance of motivation-structural rules. *Anim. Behav.* 59: 501–512.

Hallberg, K. I., Boysen S. T., and Mukobi, K. (1999). Chimpanzee food barks as referential signals: Evidence from a laboratory playback experiment [abstract]. In *22nd Meeting of the American Society of Primatologists*, New Orleans, LA.

Hauser, M. D. (1998). Functional referents and acoustic similarity: Field playback experiments with rhesus monkeys. *Anim. Behav.* 55: 1647–1658.

Hauser, M. D., and Wrangham, R. W. (1987). Manipulation of food calls in captive chimpanzees. *Folia Primatol.* 48: 207–210.

Hohman, G., and Fruth, B. (1994). Structure and use of distance calls in wild bonobos, *(Pan paniscus)*. *Int. J. Primatol.* 15: 767–782.

Hopkins, W. D., and Savage-Rumbaugh, E. S. (1991). Vocal communication as a function of differential rearing experiences in *Pan paniscus:* A preliminary report. *Int. J. Primatol.* 12: 559–583.

Jurgens, U. (1988). Central control of monkey calls. In Todt, D., Goedeking, P., and Symmes, D. (eds.), *Primate Vocal Communication*, Springer, Berlin, pp. 162–167.

Lieberman, P. (1998). *Eve Spoke: Human Language and Human Evolution*, Norton, W. W. & Co, New York.

Maeda, T., and Masataka, N. (1987). Locale-specific vocal behaviour of the tamarin *(Saguinus l. labiatus)*. *Ethology* 75: 25–30.

Marler, P. (1969). Vocalizations of wild chimpanzees. *Recent Adv. Primatol.* 1: 94–100.

Marshall, A. J., Wrangham, R. W., and Arcadi, A. C. (1999). Does learning affect the structure of vocalizations in chimpanzees? *Anim. Behav.* 58: 825–830.

Masataka, N. (1988). The response of red-chested moustached tamarins to long calls from their natal and alien populations. *Anim. Behav.* 36: 55–61.

Masataka, N., and Biben, M. (1987). Temporal rules regulating affiliative vocal exchanges of squirrel monkeys. *Behaviour* 101: 311–319.

McCowan, B., and Newman, J. D. (2000). The role of learning in chuck call recognition by young squirrel monkeys *(Saimiri sciureus)*. *Behaviour* 137: 279–300.

Meshbane, A. (1996), *Predictive Discriminant Analysis Versus Logistic Regression for Two-Group Classification Problems in Educational Settings*, Atlantic University Press, Boca Raton, FL.

Mitani, J. C., Hasegawa, T., Gros-Luis, J., Marler, P., and Byrne, R. (1992). Dialects in wild chimpanzees? *Am. J. Primatol.* 27: 233–243.

Mori, A. (1982). An ethological study on chimpanzees at the artificial feeding place in the Mahale Mountains, Tanzania, with special reference to the booming situation. *Primates* 23(1): 45–65.

Mori, A. (1983). Comparison of the communicative vocalizations and behaviors of group ranging in Eastern gorillas, chimpanzees, and pygmy chimpanzees. *Primates* 24(4): 486–500.

Newman, J. D., Smith, H. J., and Talmage-Riggs, G. (1983). Structural variability in primate vocalizations and its functional significance: An analysis of Squirrel Monkey Chuck Calls. *Folia Primatologica*. 40: 114–124.

Norcross, J. L., and Newman, J. D. (1997). Social context affects phee call production by nonreproductive common marmosets *(Callithrix jacchus)*. *Am. J. Primatol.* 43: 135–146.

Norcross, J. L., Newman, J. D., and Cofrancesco, L. M. (1999). Context and sex differences exist in the acoustic structure of phee calls by newly-paired common marmosets (*Callithrix jacchus)*. *Am. J. Primatol.* 49(2): 165–181.

Pierce, J. D. (1985). A review of attempts to condition operantly alloprimate vocalizations. *Primates* 26(2): 202–213.

Randolph, M. C., and Brooks, B. A. (1967). Conditioning of a vocal response in a chimpanzee through social reinforcement. *Folia Primatol.* 5: 70–79.

Savage-Rumbaugh, E. S. (1986). *Ape Language: From Conditioned Response to Symbol*, Columbia University Press, New York.

Savage-Rumbaugh, E. S., Murphy, J., Sevcik, R. A., Brakke, K. E., Williams, S. L., and Rumbaugh, D. M. (1993). Language comprehension in ape and child. *Monogr. Soc. Res. Child Dev.* 58: 1–256.

Sevcik, R. A., and Savage-Rumbaugh, E. S. (1994). Language comprehension and use by great apes. *Lang. Commun.* 14: 37–58.

Seyfarth, R. M., Cheney, D. L., and Marler, P. (1980). Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication. *Science* 210: 801–803.

Smith, H. J., Newman, J. D., and Symmes, D. (1982). Vocal concomitants of affiliative behavior in squirrel monkeys. In Petersen, M. R. (ed.), *Primate Communication*, Cambridge University Press, Cambridge.

Snowdon, C. T., and Cleveland, J. (1984). "Conversations" among pygmy marmosets. *Am. J. Primatol.* 7(1): 15–20.

Struhsaker, T. T. (1967). Auditory communication among vervet monkeys *(Cercopithecus aethiops)*. In Altman, S. A. (ed.), *Social Communication Among Primates*, University of Chicage Press, Chicago. 281–324.

Sugiura, H. (1993). Temporal and acoustic correlates in vocal exchange of coo calls in Japanese macaques. *Behaviour* 124(3/4): 207–225.

Sugiura, H. (1998). Matching of acoustic features during the vocal exchange of coo calls by Japanese macaques. *Anim. Behav.* 55(3): 673–687.

Sutton, D., Larson, C., Taylor, E. M., and Lindemen, R. C. (1973). Vocalization in rhesus monkeys: Conditionability. *Brain Res.* 52: 225–231.

Sutton, D., Samson, H. H., and Larson, C. R. (1978). Brain mechanisms in learned phonation of *Macaca mulatta*. In Chivers, D. J., and Herbert, J.(eds.), *Recent Advances in Primatology: Vol. 1: Behavior*, Academic Press, London, pp. 769–784.

Sutton, D., Trachy, R. E., and Lindemen, R. C. (1981). Vocal and nonvocal discriminative performance in monkeys. *Brain Lang.* 14: 93–105.

Symmes, D., and Biben, M. (1988). Conversational vocal exchanges in squirrel monkeys. In Todt, D., Goedeking, P., and Symmes, D. (eds.), *Primate Vocal Communication*, Springer, Berlin.

Tabachnick, B. G., and Fidell, L. S. (1996). *Using multivariate statistics*, 3rd edition, Harper Collins, New York.

van Hoof, J. A. R. A. M. (1972). A comparative approach to the phylogeny of laughter and smiling. In Hinde, R. A. (ed.), *Nonverbal Communication* Cambridge University Press, Cambridge, UK. pp. 12–53.

Zuberbühler, K. (2000). Interspecies semantic communication in two forest primates. *Proc. R. Soc. Lond. B* 267: 713–718.

Zuberbuhler, K., Cheney, D. L., and Seyfarth, R. M. (1999). Conceptual semantics in a nonhuman primate. *J. Comp. Psychol.* 113: 33–42.