Influence of emotion and focus location on prosody in matched statements and questions

Marc D. Pell[a)]
School of Communication Sciences and Disorders, McGill University, 1266 Pine Avenue West, Montréal, Québec H3G 1A8, Canada

(Received 5 May 2000; accepted for publication 4 January 2001)

Preliminary data were collected on how emotional qualities of the voice (sad, happy, angry) influence the acoustic underpinnings of neutral sentences varying in location of intra-sentential focus (initial, final, no) and utterance ''modality'' (statement, question). Short (six syllable) and long (ten syllable) utterances exhibiting varying combinations of emotion, focus, and modality characteristics were analyzed for eight elderly speakers following administration of a controlled elicitation paradigm (story completion) and a speaker evaluation procedure. Duration and fundamental frequency ($f0$) parameters of recordings were scrutinized for ''keyword'' vowels within each token and for whole utterances. Results generally re-affirmed past accounts of how duration and $f0$ are encoded on key content words to mark linguistic focus in affectively neutral statements and questions for English. Acoustic data on three ''global'' parameters of the stimuli (speech rate, mean $f0$, $f0$ range) were also largely supportive of previous descriptions of how happy, sad, angry, and neutral utterances are differentiated in the speech signal. Important interactions between emotional and linguistic properties of the utterances emerged which were predominantly (although not exclusively) tied to the modulation of $f0$; speakers were notably constrained in conditions which required them to manipulate $f0$ parameters to express emotional and nonemotional intentions conjointly. Sentence length also had a meaningful impact on some of the measures gathered. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1352088]

PACS numbers: 43.70.Fq [AL]

I. INTRODUCTION

Recent decades have yielded numerous insights into the acoustic properties of human discourse, including its suprasegmental or ''prosodic'' attributes. For the most part, the purview of these studies has revolved around presumed linguistic or emotional features of the signals under scrutiny, limiting acoustic data on prosody to a description of one of these two operational contexts. Meager attention has been accorded to how linguistic and emotional attributes of spoken utterances *interact* to influence the acoustic form of speech. The present study addresses this issue, contributing to an initial understanding of the *conjoint* influences of emotive and linguistic processes on the acoustic form of spoken utterances in English. The current design also permitted important corroborative data on previously described prosodic phenomena in the acoustic literature (e.g., Eady and Cooper, 1986; Williams and Stevens, 1972) for a somewhat older group of speakers.

A. Acoustic investigations of emotional or linguistic prosody

Many researchers have sought to characterize external acoustic modifications linked to a speaker's emotional disposition and associated processes (e.g., Banse and Scherer, 1996; Frick, 1985; Sobin and Alpert, 1999). A sizable, independent literature has accumulated on how speakers exploit suprasegmental parameters to assign linguistic meaning in discourse (e.g., Bolinger, 1955; Cooper *et al.*, 1985; Ladd, 1996). In both areas, most studies have concentrated on the operation of fundamental frequency ($f0$) and duration in the transmission of prosodic meaning (Murray and Arnott, 1993), although important data on vocal intensity and ''voice quality'' have also been gathered (Bachorowski and Owren, 1995; Cummings and Clements, 1995; Denes, 1959; Lieberman and Michaels, 1962; Scherer, 1974; Sobin and Alpert, 1999; Turk and Sawusch, 1996).

Acoustic properties of *affective* vocalizations, like other discernable cues to emotional processes (e.g., facial, gestural, and body movements), are shaped by patterned neurophysiological responses coupled with a speaker's attempt to regulate their response due to social-contextual factors (Ekman *et al.*, 1983; Scherer, 1986; Siegman *et al.*, 1992). The speaker's physiological arousal during communication, and to a lesser extent, the relative pleasantness (*valence*) of the emotion being experienced, exert a particularly strong influence on the external configuration of emotion-related cues when ''internal'' factors are considered (Davitz, 1964; Huttar, 1968; Ladd *et al.*, 1985; Murray and Arnott, 1993; Pakosz, 1983; Scherer, 1986; Siegman and Boyle, 1993; Uldall, 1960). Given their origin in relatively widespread alterations in a speaker's emotional, physiological, and cognitive status, specifying the acoustic cues to particular types of emotional responses has proven a challenge (see Bachorowski, 1999, for a discussion). Of particular note, collecting naturalistic exemplars of vocal emotion in speech while controlling for linguistic/segmental variables and the recording quality of obtained samples has proven complicated, although some

a)Electronic mail: marc.pell@mcgill.ca

spontaneous recordings of emotion have been analyzed (Streeter *et al*., 1983; Williams and Stevens, 1972).

Despite methodological obstacles—and the inherent compromises associated with analysis of both spontaneous and ''simulated'' emotions—research on vocal emotion expression has identified several acoustic parameters associated with purported ''basic'' emotion states (Ekman, 1992; Izard, 1977). To summarize this relationship for the emotions typically investigated, happy, angry, and fearful speech tend to display a higher mean $f0$ and amplitude, with greater $f0$ and amplitude variation, than emotionally neutral utterances; sad utterances exhibit the opposite pattern relative to neutral utterances (lower mean $f0$ and amplitude, minimal $f0$/amplitude variation) (Banse and Scherer, 1996; Breitenstein *et al*., in press; Davitz, 1964; Fairbanks and Pronovost, 1939; Huttar, 1968; Sobin and Alpert, 1999; Williams and Stevens, 1972). Speech rate also differentiates among emotional modes, neutral and angry utterances demonstrating an accelerated rate of delivery relative to sad utterances, for example (Breitenstein *et al*., in press; Fairbanks and Hoaglin, 1941; Johnson *et al*., 1986; Williams and Stevens, 1972). Many of these tendencies in how ''discrete'' emotions are differentiated acoustically are captured by a hypothetical model of affect expression formulated by Scherer (1986).

Not all suprasegmental forms reflect the outward manifestation of emotional processes and associated responses. Speakers routinely employ prosody to encode *linguistic* messages based on conventionalized knowledge about their language system and its sociolinguistic applications (Bolinger, 1978). Unlike emotional attributes, acoustic correlates of linguistic prosody operate at circumscribed levels of linguistic representation such as the syllable, word, or utterance, directing the manner in which these phenomena have been studied. For example, research on linguistic prosody in English has illustrated (among others) that ''stressed'' syllables are associated with an elevated peak $f0$ and amplitude, increased $f0$ and amplitude variation, and an elongated vocalic segment when compared to unaccented syllables produced in the same linguistic environment (Brown and McGlone, 1974; Cooper and Sorensen, 1981; Klatt, 1976; Lea, 1977; McClean and Tiffany, 1973; Morton and Jassem, 1965). Words assigned linguistic focus within an utterance display a similar constellation of phonetic features as stressed syllables or words (Eefting, 1990; Ferreira, 1993; Folkins *et al*., 1975; Weismer and Ingrisano, 1979), although more extensive modifications are witnessed in the case of focus (Bolinger, 1958; Fry, 1955, 1958).

Speech acts, or utterances varying in their linguistic-pragmatic intention with respect to the listener (e.g., statements/questions), have also been characterized acoustically; this research points to the critical importance of $f0$ parameters in the terminal portion of the utterance for many languages (Bolinger, 1978; Lieberman, 1967; Ohala, 1983). However, additional points in the $f0$ contour may also differentiate the declarative and interrogative ''modality'' (Hadding-Koch and Studdert-Kennedy, 1964; Majewski and Blasdell, 1968; O'Shaughnessy, 1979; Studdert-Kennedy and Hadding, 1973). A series of investigations undertaken by Cooper, Eady, and colleagues (Cooper *et al*., 1985; Eady and Cooper, 1986; Eady *et al*., 1986) furnished important data on how linguistic applications of prosody operating at the word- and utterance-levels *interact* in the speech signal for English. Utilizing a structured elicitation paradigm with a limited number of speakers, these authors analyzed prominent duration and $f0$ attributes of sentential ''key words'' according to the modality of the utterance (statement, question) and the location of contrastive focus within the stimulus (initial, medial, final key word). These reports corroborated prominent duration and $f0$ cues associated with linguistic focus and speech acts summarized above, and were instrumental in exemplifying the *interaction* of these linguistic processes on localized acoustic parameters of spoken utterances, as well as the overall shape of the intonation contour (see Eady and Cooper, 1986). The issue of utterance *length* and its potential impact on the acoustic properties of focus in statements and questions was also raised by the researchers (Eady and Cooper, 1986; Eady *et al*., 1986), although length, focus, and modality effects on prosody implementation were not examined directly by the researchers.

B. Intersection of emotional and linguistic prosody in English

Thus increasingly sophisticated views of how speech prosody operates in emotional or nonemotional contexts are tempered by a lack of controlled experimentation looking at the *intersection* between emotional and linguistic constructs on the acoustic form of utterances (cf. Cosmides, 1983; McRoberts *et al*., 1995; Ross *et al*., 1986). Understanding how emotional and linguistic prosody interact in the speech signal will supply vital information on human vocal communication with its concurrent demands for affective and propositional signaling, despite access to a limited and overlapping set of acoustic features available to express these different intentions through speech. The present investigation adopts methods employed by Cooper, Eady, and colleagues in their exploration of linguistic prosody in English, adapting this procedure to evaluate acoustic dimensions of focused words in matched statements and questions in three simulated *emotional* contexts (happy, angry, sad). New data regarding the interaction of emotional and linguistic factors on suprasegmental parameters were further extended to ''short'' and ''long'' utterances elicited from a single group of encoders (Eady and Cooper, 1986; Weismer and Ingrisano, 1979). This design will allow novel insights into factors that affect the acoustic realization of linguistic meanings through prosody and how these representations are accommodated by ''prototypical'' emotional qualities of the voice during speech in a manner capable of advancing future research in this area.

II. METHODS

A. Subjects

Five female and five male normally aging adults (mean = 66.1 years, range = 59–72) volunteered for the study; these individuals also served as a control group in studies examining speech production characteristics following unilateral brain damage (Pell, 1999a, b, c). Subjects were

native speakers of Canadian English with at least eight years of formal education and no reported history of speech, language, or neurological disturbance. All participants displayed good hearing following a pure-tone air conduction screening (entry criteria: 30 dB HL at 0.5, 1, and 2 kHz, for the better ear).

B. Materials

Stimuli were two six-syllable (''*short*'') and two ten-syllable (''*long*'') English utterances matched for stress assignment at designated ''keyword'' positions (italicized):

Short								
1.	*Barry*	*took*	the	*sailboat*				
2.	*Mary*	*sold*	the	*teapot*				
Keyword:	FOC1	INT1		FOC2				
Long								
1.	*Barry*	*took*	the	*sailboat*	for	the	*weekend*	
2.	*Mary*	*sold*	the	*teapot*	for	a	*dollar*	
	FOC1	INT1		INT2			FOC2	

Long and short items were differentiated by a terminal prepositional phrase and by the number of keyword sites in the stimulus. Both short and long stimuli contained two keyword ''candidates'' for focus realization in particular contexts (FOC—underlined above). However, short stimuli contained only one rather than two ''intervening'' keywords (INT) where acoustic measures were later derived. Stimuli could be produced in three focus contexts: no focus, sentence-initial focus (focus on FOC1), or sentence-final focus (focus on FOC2). Each test item could also be intoned as a statement or yes–no question without subject-auxiliary inversion in each focus context. Finally, each item was conducive to four emotional interpretations (neutral, sad, happy, angry) which could not be inferred from the lexical-semantic content, but only from the prosody.[1] An exhaustive combination of focus, modality, and emotion features yielded 24 prosodically unique exemplars of each item or 96 productions per speaker (2 items×2 lengths×3 focus contexts×2 modalities×4 emotions).

A recorded vignette preceded each trial to elicit tokens conveying specific combinations of prosodic attributes (Cooper *et al.*, 1985; Eady and Cooper, 1986; Ryalls *et al.*, 1994). Vignettes consisted of a question or short passage that biased the target response; half of the recordings biased a response with a statement intonation and half with a question intonation. Differences in focus were rendered by modifying the vignette to place information to be used contrastively as ''new'' or unresolved within the situational context (e.g., Eady and Cooper, 1986). For example, the scenario preceding a neutral, interrogative reading of *Mary sold the teapot* with final focus (i.e., [final focus, question, neutral]) was the following:

> *You are holding a garage sale at your house with the help of some friends. After the sale, you notice that Mary sold many of the articles on her table. To find out whether the teapot was one of the items that was sold, you turn to another friend and ask*:
> [Mary sold the teapot]

To elicit emotional renditions of target utterances, vignettes provided explicit situational cues consistent with the target emotion. For example, to bias a sad interpretation of the passage above, it was explained to the listener that the teapot had strong sentimental value and had been included in the sale by accident; the vignette then terminated with a direct prompt to respond in the target emotion (e.g., *you say/ask sadly...*). Vignettes were recorded in a soundproof chamber by a male speaker who was coached to produce the passages in a neutral, ''reporting'' tone. The tape containing the 96 scenarios was edited to randomize trials for order of presentation and to insert a 5-s interstimulus pause between vignettes for subjects to execute a response.

C. Procedure

Subjects were tested individually during two 30-min sessions to limit fatigue and possible inattention to pre-recorded vignettes. Subjects were seated comfortably with a directional microphone (Sony ECM-909) 20 cm in front of their mouths and stimulus cards placed on a table in front of them. Participants were encouraged to pay close attention to the vignettes (presented free-field) and to complete the story by producing the sentence on the card in front of them (cards also contained visual information reinforcing target prosodic features, such as bold italics to signify focused words). Repetition of subjects' responses occurred rarely but was permitted in the event of reading errors, dysfluencies, or subject-initiated ''corrections''; for such items, the final production was considered for further analysis. Five practice trials acquainted subjects with the experimental procedure. All responses were recorded onto digital audio tape and re-digitized using the BLISS speech analysis system (Mertus, 1989) at a sampling rate of 20 kHz, with a 9 kHz low-pass filter setting and 12-bit quantization.

D. Speaker evaluation

To accurately portray the conjoint influences of specific emotion and focus characteristics on the acoustic form of utterances, it was imperative to limit acoustic analyses to speakers who adequately encoded these target meanings in their productions. This selection process was critical to the aims of the current investigation by mitigating the potential effects of encoders who did not accommodate well to the elicitation procedure on the acoustic findings (also Eady and Cooper, 1986). Speaker evaluation criteria took the form of listener judgments about the emotion expressed or about the word being focused for a subset of each speaker's utterances (the 12 declarative readings of ''*Mary sold the teapot*''). These stimuli elicited from each of the ten encoders were randomized within a single task and presented to ten listeners twice. On one occasion, listeners identified where the heard focus within the utterance (initial word, final word, none). On a separate occasion listeners identified the speakers' emotional tone for these tokens (neutral, sad, happy, angry). Results confirmed anticipated differences among encoders in how reliably they transmitted both types of intentions to listeners (e.g., Sobin and Alpert, 1999, for data on emotional

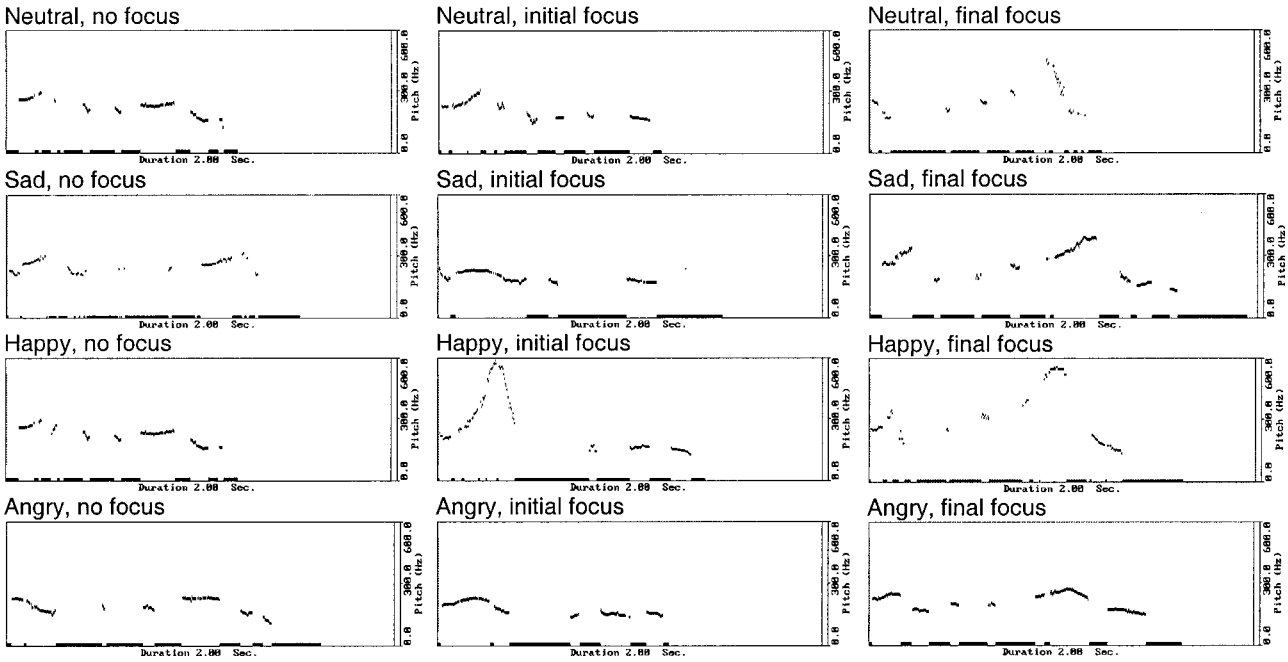

FIG. 1. *F*0 contours representing 12 distinct productions of *Barry took the sailboat* by one of the ''proficient'' female speakers.

judgments); correct listener identifications of focus and emotion targets in productions elicited from the ten encoders was 78.6% (±13.7) and 67.2% (±14.6), respectively.[2] Based on the overall group ratings, two male speakers who fell below *both* the focus and emotion condition means were excluded from further acoustic/statistical consideration. It was observed that three female encoders were especially adept at communicating linguistic and emotional meanings in tandem, yielding correct perceptual judgments of their speech exceeding 80% in both emotion and focus conditions. A sample of utterances produced by one of these ''proficient'' encoders is showcased in Fig. 1.

E. Acoustic/statistical analyses

Duration and $f0$ measures were extracted at critical ''keyword'' sites in each utterance (Eady and Cooper, 1986) and for each stimulus as a whole (e.g., Ladd *et al.*, 1985; Williams and Stevens, 1972). For *keyword measures*, duration (in ms) was computed for the full vowel (transition and steady state) on the stressed syllable of each content word (e.g., Ma-ry, sold, tea-pot, dol-lar). The corresponding $f0$ of these vowels—roughly the ''topline'' of the intonation contour (Cooper *et al.*, 1985; Eady and Cooper, 1986)—was also computed, as was the mean $f0$ of the terminal 150 ms (T) of the intonation contour (Lieberman, 1967). For both vowel and terminal measures, $f0$ was extracted via an autocorrelation algorithm and then five contiguous pulses were isolated visually at the center of the target constituent by placing cursors at zero crossings and the inverted period in Hz was averaged (Behrens, 1988; Ouellette and Baum, 1993). For whole utterance measures, three parameters were isolated for analysis: *speech rate*, or the number of syllables divided by the total sentence duration in ms; *mean* $f0$, the average of all keyword $f0$ values within a token including the terminal measure; and $f0$ *range*, the calculated difference between the maximum and minimum $f0$ value derived at keyword sites including the terminal point.

Acoustic measures were normalized for each speaker prior to statistical analysis. Keyword vowel durations were divided by the corresponding utterance duration to adjust for inter-speaker differences in speech rate. Keyword and utterance $f0$ measures were normalized for inter-speaker differences in $f0$ range by applying the following z score transform to each value: $f0i=(f0i-f0\text{mean})/s$, where $f0i$ is the observed $f0$ value, $f0$ mean is the mean $f0$ across all utterances produced by the speaker (both short and long), and s is the standard deviation (Gandour *et al.*, 1995; Rose, 1987). *F*0 range was normalized for inter-subject differences in mean $f0$ by dividing the range by the corresponding utterance mean for each speaker. Statistical analyses were performed separately on each acoustic parameter, and for keyword measures, at each keyword position [FOC1, INT1, INT2 (long only), FOC2, *T*]. Data gathered for the two ''short'' and the two ''long'' items were first collapsed in each condition. A total of eight $2\times3\times2\times4$ repeated measures ANOVAs were performed, with LENGTH (short, long), FOCUS (no, initial, final), MODALITY (statement, question), and EMOTION (neutral, sad, happy, angry) serving as within-subjects fixed variables. Geisser–Greenhouse $\tilde{\varepsilon}$ adjusted critical values determined the significance of tests involving more than 1 df on within-subjects factors (Max and Onghena, 1999) and a conservative p of 0.01 was adopted to focus discussion on the most robust patterns in the acoustic data. *Post hoc* comparisons were conducted using Tukey's HSD procedure ($p<0.01$), where appropriate.

III. RESULTS

Results first examine how localized ''keyword'' measures of duration and $f0$ are associated with focus realization

TABLE I. Mean duration[a] of sentence-initial (FOC1) and sentence-final (FOC2) keyword vowels produced in four emotional tones as a function of utterance length, modality and focus context. Shaded cells indicate keywords when focus was realized in each condition.

			NEUTRAL Keyword position		**SAD** Keyword position		**HAPPY** Keyword position		**ANGRY** Keyword position	
Length	Modality	Focus	FOC1	FOC2	FOC1	FOC2	FOC1	FOC2	FOC1	FOC2
SHORT	(.)	no	134	106	127	99	118	113	122	107
		initial	150	92	159	89	153	100	159	86
		final	113	118	118	113	113	124	113	121
	(?)	no	116	102	127	102	124	108	115	106
		initial	164	104	160	96	147	101	150	101
		final	120	109	116	103	115	115	117	114
LONG	(.)	no	88	78	87	83	85	89	83	86
		initial	112	69	109	74	106	79	106	78
		final	77	89	79	92	76	94	80	91
	(?)	no	83	82	86	82	82	88	85	88
		initial	109	80	112	79	105	86	113	79
		final	78	85	76	92	77	92	76	87

[a]Vowel durations (in ms) were adjusted for inter-speaker differences in speaking rate and then multiplied by 1000.

in declarative utterances, in matched interrogative utterances, in short versus long stimuli, and in stimuli expressed in different emotional modes. Data reflecting ''global'' dimensions of the same utterances are subsequently examined to elaborate on how emotion influences acoustic properties of spoken utterances that encode specific combinations of the prosodic features.

A. Keyword measures

Table I shows the duration of sentence-initial (FOC1) and sentence-final (FOC2) keyword vowels and Table II displays the corresponding $f0$ of these vowels, of ''intervening'' content words (INT1, INT2), and of the utterance terminal (*T*). All values are an average of two items elicited in each condition, collapsed (for expository purposes) across eight speakers (Table I) or across female or male speakers (Table II).

1. Focus realization in neutral statements: FOCUS main effect

Duration and $f0$ attributes of contrastive focus in English have been established by several investigators (e.g., Cooper *et al.*, 1985; Eady and Cooper, 1986; Weismer and Ingrisano, 1979) and are largely corroborated here. Focus realization corresponded to a lengthening of focused versus unfocused target vowels at both FOC1 and FOC2 [FOCUS: $F_{\mathrm{FOC1}}(2,14)=53.94$, $p<0.001$; $F_{\mathrm{FOC2}}(2,14)=28.22$, $p<0.001$]. Focused vowels simultaneously exhibited higher $f0$ peaks than corresponding unfocused vowels at each of these two positions [FOCUS: $F_{\mathrm{FOC1}}(2,14)=11.86$, $p<0.01$; $F_{\mathrm{FOC2}}(2,14)=27.12$, $p<0.001$]. These well-accepted parameters likely represent central features of how ''new'' information is highlighted for the listener in many spoken languages (Bolinger, 1986; Fowler and Housum, 1987).

Inspection of the data further looked at how the context for focus realization (initial, final, no focus) influenced the acoustic form of specific keywords in statements. For duration, signaling focus on the initial word of utterances (FOC1) had a significant effect on the acoustic properties of *final* keywords (FOC2); as illustrated in Fig. 2, FOC2 vowels spoken ''post-focus'' (initial-focus condition) were systematically shorter than when FOC2 was itself focused (final focus), and importantly, than in speaking conditions where *no* contrastive focus was realized in the utterance. Post-focal reduction of the $f0$ of FOC2 was also witnessed for the stimuli, yielding significant decrements in the peak $f0$ values of FOC2 in the context of ''post'' (initial) focus [see Fig. 3(a)]. These patterns indicate maximal acoustic differentiation of FOC2 as an index of the three focus contexts for both duration and $f0$. In contrast, a reciprocal process of acoustically ''de-accentuating'' FOC1 vowels in anticipation of focus at FOC2 (i.e., pre-focus effects) beyond those cues expected in utterances without focus altogether was not a statistically reliable attribute of the stimuli for either duration or $f0$.

2. Focus realization in statements versus questions: Effects of MODALITY

Stimuli formulated as a question rather than statement led to expected changes in the $f0$ of all keyword values following sentence-initial words (FOC1), consistent with a fall/nonfall pattern for statements and questions respectively (e.g., Eady and Cooper, 1986; Studdert-Kennedy and Hadding, 1973). This diverging trend was significantly distinctive within the final 150 ms of the $f0$ contour where statements displayed a lower terminal $f0$ than questions in all conditions [MODALITY: $F_T(1,7)=88.41$, $p<0.001$]. The *conjoint* influences of linguistic modality and focus context on the $f0$ contour are illustrated in Fig. 3 and were most notable when focus was realized at FOC1; this led to an abrupt lowering (statement) or elevation (question) of $f0$ by the speaker on immediate post-focus content words [MODALITY×FOCUS: $F_{\mathrm{INT1}}(2,14)=14.34$, $p<0.01$]. This marked $f0$ contrast was maintained by speakers at all remaining keyword locations in utterances spoken in the initial-focus context [MODALITY×FOCUS: $F_{\mathrm{INT2}}(2,14)=33.54$, $p<0.0001$; $F_{\mathrm{FOC2}}(2,14)=22.63$, $p=0.001$] (Eady and Cooper, 1986). Rendering matched utterances in con-

TABLE II. Mean *f*0 of vowels produced in sentence positions of potential focus (FOC), for intervening content words (INT), and for the ''terminal'' (*T*) portion of utterances spoken in four emotional tones (A–D). Data are presented separately for female ($n=5$) and male ($n=3$) speakers. Shaded cells indicate keywords where focus was realized in each condition as a function of utterance length and modality.

			SHORT Keyword Position				**LONG** Keyword Position				
Modality	Focus	Gender	FOC1	INT1	FOC2	*T*	FOC1	INT1	INT2	FOC2	*T*
A-**NEUTRAL**											
(.)	no	female	195	172	202	158	217	194	190	194	150
		male	113	115	162	74	116	118	130	154	97
	initial	female	242	143	136	111	263	157	153	150	132
		male	179	100	91	74	210	104	100	85	78
	final	female	213	186	306	125	211	193	189	208	129
		male	117	136	233	89	126	117	135	184	94
(?)	no	female	191	236	211	349	187	233	210	169	341
		male	134	119	147	259	120	126	138	121	234
	initial	female	183	280	264	369	184	260	264	241	348
		male	147	164	171	252	149	157	173	159	238
	final	female	210	214	214	331	201	215	204	171	316
		male	132	121	181	275	126	127	128	149	262
B-**SAD**											
(.)	no	female	208	187	192	134	203	179	192	189	159
		male	137	127	140	75	129	111	118	131	89
	initial	female	203	154	157	140	220	166	162	151	150
		male	139	104	90	72	155	107	120	109	84
	final	female	223	189	223	140	217	181	183	200	130
		male	124	118	125	83	133	110	125	159	111
(?)	no	female	190	225	204	301	214	203	189	160	294
		male	130	133	152	241	123	120	131	129	226
	initial	female	185	225	230	314	190	218	207	178	264
		male	128	158	166	242	142	141	169	151	225
	final	female	187	201	191	307	205	191	192	176	273
		male	130	110	160	234	123	115	131	146	241
C-**HAPPY**											
(.)	no	female	275	208	311	143	307	211	260	298	141
		male	174	128	253	89	197	139	155	214	102
	initial	female	367	180	168	129	341	181	178	207	111
		male	229	140	129	92	240	131	144	140	96
	final	female	287	224	409	144	257	193	212	335	154
		male	154	132	245	85	151	149	160	234	98
(?)	no	female	261	257	258	430	236	235	224	218	360
		male	134	162	172	293	137	125	135	145	220
	initial	female	221	258	272	414	244	255	256	238	370
		male	153	160	200	251	181	155	188	182	241
	final	female	219	213	239	413	239	208	207	222	373
		male	124	135	178	323	132	134	132	134	286
D-**ANGRY**											
(.)	no	female	222	190	261	120	216	187	214	222	104
		male	154	158	134	86	149	142	161	166	87
	initial	female	235	152	158	113	244	184	169	176	153
		male	160	102	100	86	166	124	121	107	84
	final	female	232	185	277	115	217	187	209	242	153
		male	141	133	193	90	141	138	146	181	93
(?)	no	female	238	225	247	311	239	212	232	213	318
		male	143	144	190	280	152	120	157	150	261
	initial	female	219	238	239	330	232	235	250	221	321
		male	151	163	187	264	166	180	189	166	247
	final	female	201	193	231	376	225	197	212	212	347
		male	115	118	178	277	147	139	144	166	265

texts of no-focus or final-focus was associated with less pronounced modifications to the global shape of statement/question contours (review Fig. 3).

There was also evidence that sentences formulated as a question affected how isolated target vowels at FOC1 and FOC2 were encoded to express the context for focus, elaborating patterns reported in the preceding section for statements only. As may be seen in Figs. 2 and 3 for duration and

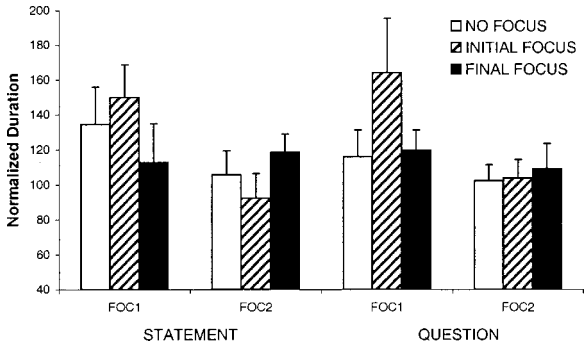

FIG. 2. Duration of sentence-initial (FOC1) and sentence-final (FOC2) vowels produced in short statements and questions in three focus contexts (averaged across eight speakers).

$f0$, respectively, stimuli posed as questions contained a more limited set of acoustic features that biased the focus context of the utterance than statements when measures were restricted to FOC1 and FOC2. In the case of duration (Fig. 2), questions were associated with focus-related vowel lengthening at both FOC1 and FOC2 as described earlier for statements; however, in questions, there was no significant evidence of post-focal vowel reduction at FOC2 in the initial-focus context beyond that associated with no focus [MODALITY×FOCUS: $F_{\mathrm{FOC2}}(2,14)=13.84$, $p=0.001$]. In the case of $f0$ (Fig. 3), formulating stimuli as a question had an even more pronounced impact on localized cues to focus at FOC1 and FOC2; unlike statements where reliable $f0$ differences characterized focused and unfocused vowels at each of these keyword positions, questions contained *no* statistically significant $f0$ differences indicative of focus context when measures were limited to FOC1 and FOC2

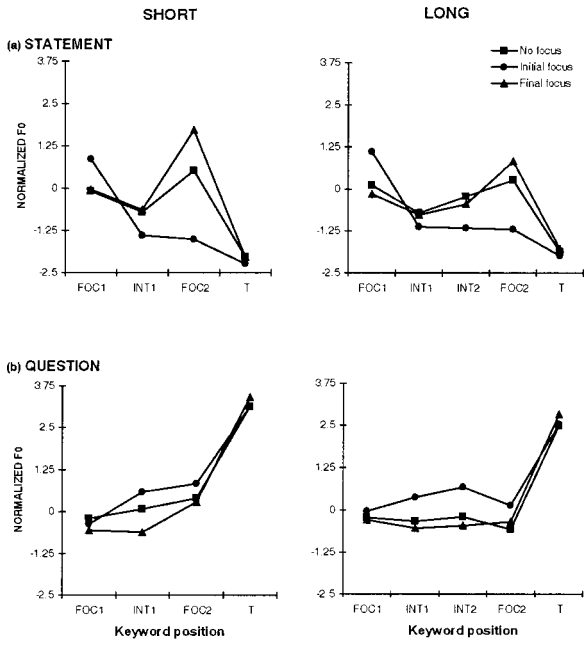

FIG. 3. Fundamental frequency of initial (FOC1), intervening (INT1, INT2), final (FOC2), and terminal (*T*) keyword vowels produced in short and long sentences as a (a) statement and (b) question (averaged across eight speakers and four emotions).

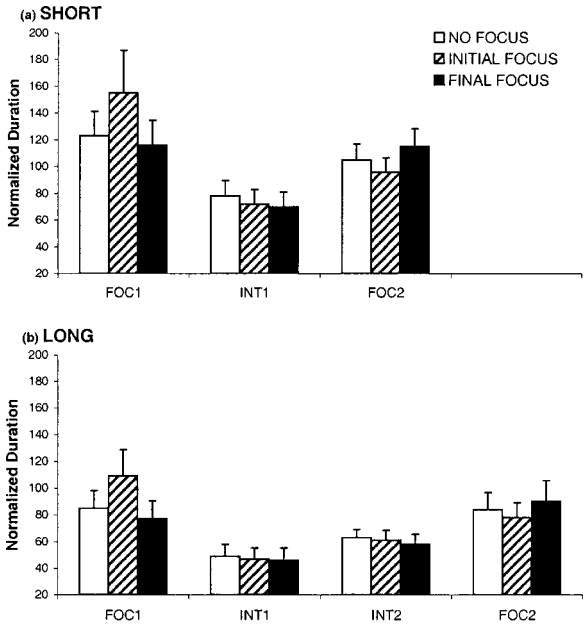

FIG. 4. Duration of initial (FOC1), intervening (INT1, INT2), and final (FOC2) keyword vowels spoken in three focus contexts for (a) short and (b) long stimuli (averaged across eight speakers and two modalities).

[MODALITY×FOCUS: $F_{\mathrm{FOC1}}(2,14)=12.57$, $p<0.01$; $F_{\mathrm{FOC2}}(2,14)=22.63$, $p=0.001$].

3. Focus realization in short versus long utterances: Effects of LENGTH

Producing ten—as opposed to six—syllable utterances did not alter the relative pattern of $f0$ cues implemented at FOC1 or FOC2 to signal different focus contexts. Only the endpoint of the $f0$ contour was influenced significantly by length, revealing differences in terminal measures for questions (''short'' terminated with a higher $f0$ than ''long'') but not for statements [LENGTH×MODALITY: $F_T(1,7)=24.59$, $p=0.002$] (review Fig. 3).

Figure 4 presents a graphic comparison of keyword duration values in short and long stimuli averaged across modality. At FOC2, it can be seen that duration parameters corresponding to the three focus contexts were not maximally distinctive in long stimuli [Fig. 4(b)] as they were in short stimuli [Fig. 4(a)], although obtained cues were in the expected direction (for long: final-focus=no-focus >initial-focus) [LENGTH×FOCUS: $F_{\mathrm{FOC2}}(2,14)=6.54$, $p=0.01$]. Of perhaps greater importance, temporal attributes of ''intervening'' words in the stimuli played a key role in marking focus interpretation for long utterances. Recall that long stimuli permitted two vowel measurements between experimentally designated focus positions (e.g., *sold* and *teapot* in *Mary sold the teapot for a dollar*). When focus was realized at FOC2 (sentence-final position), speakers systematically reduced the duration of the intervening content word directly preceding FOC2 (i.e., INT2) in this one context [FOCUS: $F_{\mathrm{INT2}}(2,14)=9.82$, $p<0.01$]. Interestingly, pre-focus anticipatory markers in long utterances were more pronounced at INT2 in questions (no-focus=initial-focus >finalfocus) than in statements (no-focus>final-focus =initial-focus) [MODALITY×FOCUS: $F_{\mathrm{INT2}}(2,14)=7.29$,

$p<0.01$]. This difference may reveal subtle mechanisms by which speakers attempt to compensate for apparent $f0$ constraints in signaling focus in interrogative utterances noted above. Focus-related alterations in the duration of INT1 (e.g., *sold*) were never observed for either short or long stimuli ($p>0.01$) (Weismer and Ingrisano, 1979).

4. Focus realization in emotional utterances: Effects of EMOTION

Documenting the impact of emotion on acoustic features reported above constituted the primary aim of the current investigation. First, as expected, varying emotional qualities of the stimuli was associated with acoustic changes that were detectable at isolated keyword locations but unrelated to other specifications of the stimuli (such as focus). These main effects were predominantly tied to the operation of $f0$, were significant at all keyword locations [EMOTION: $F_{\mathrm{FOC1}}(3,21)=34.12$, $p<0.001$; $F_{\mathrm{INT1}}(3,21)=6.85$, $p=0.01$; $F_{\mathrm{INT2}}(3,21)=9.03$, $p<0.01$; $F_{\mathrm{FOC2}}(3,21)=30.26$, $p<0.001$], and will be discussed in a later section on mean $f0$ and $f0$ variation at the *sentence* level. Of singular note here, emotion marginally influenced the terminal point of utterances overall, with happy contours demonstrating a higher endpoint than sad contours [EMOTION: $F_T(3,21)=4.92$, $p=0.017$]. Terminal $f0$ values of angry or neutral stimuli did not significantly differ in either case from those of happy or sad stimuli.

Manipulating emotional attributes of experimental stimuli had a strong influence on patterns of $f0$ implemented at FOC1 and FOC2 to mark focus context, which were both dependent on the linguistic modality of the utterance. At FOC1, significant interactions emerged for EMOTION×FOCUS [$F_{\mathrm{FOC1}}(6,42)=7.98$, $p=0.001$], EMOTION×MODALITY [$F_{\mathrm{FOC1}}(3,21)=10.30$, $p<0.01$] and MODALITY×FOCUS (reported above). Analysis of FOC2 vowels yielded a three-way interaction of EMOTION×MODALITY×FOCUS [$F_{\mathrm{FOC2}}(6,42)=6.29$, $p<0.01$] which was marginally significant at FOC1 ($p=0.08$). An overview of the interplay among emotion, focus, and modality on $f0$ characteristics of FOC1 and FOC2 vowels is illustrated in Fig. 5 for short stimuli only.

In summarizing the three-way relationship, it is first worth stating that differences in emotion did not lead to systematic alterations in how $f0$ was implemented at FOC1 or FOC2 to communicate the context for focus; rather, the interplay of focus and emotion was mostly attributable to differences among the four levels of emotion that were not entirely uniform when each of the focus contexts was examined separately. In general, the $f0$ of FOC1 and FOC2 produced in a happy tone consistently surpassed that spoken in a sad or neutral tone. The same keywords expressed in angry stimuli exhibited a higher $f0$ than those in sad stimuli (where $f0$ tended to be lowest) but this comparison was significant in fewer focus contexts. Importantly, these qualitative patterns were strongly dependent on the linguistic modality of the utterance and, when inspected carefully, were only true of statements. For questions, rather, there were no statistically reliable differences among the four emotions in any of the three focus contexts when measures were restricted to the purview of FOC1 and FOC2 [compare Figs. 5(a)–(c)].

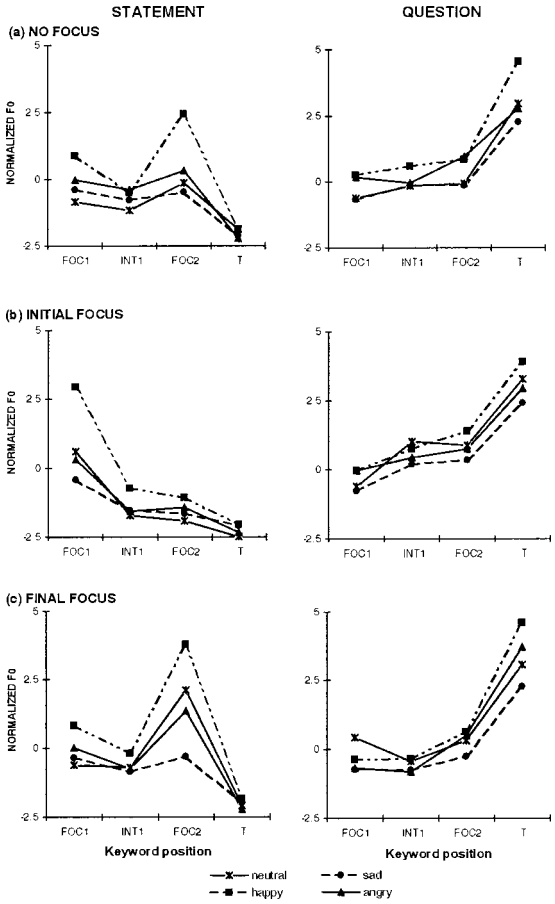

FIG. 5. Influence of emotion on the $f0$ of initial (FOC1), intervening (INT1), final (FOC2) and terminal (*T*) keyword vowels in short statements and questions in the (a) no focus, (b) initial focus, and (c) final focus context (averaged across eight speakers).

Varying emotional qualities of speech had comparatively little effect on *duration* attributes of keywords, affecting only FOC2 and in a manner unrelated to focus or any other prosodic manipulation [EMOTION: $F_{\mathrm{FOC2}}(3,21)=6.71$, $p<0.01$]. Speakers systematically prolonged final words in utterances expressed in a happy tone relative to those expressed in an angry, sad, or neutral tone in this position (which did not differ in any case). This effect did not interfere with temporal parameters of FOC2 vowels that signalled focus realization in the utterance, including post-focus vowel reduction in the production of neutral statements.

B. Utterance measures

Interplay of emotion and linguistic variables on global acoustic properties

Acoustic properties of whole utterances were analyzed to further illuminate potential emotion effects on target productions for three measures: speech rate, mean $f0$, and $f0$ range. For the speech rate data, significant effects were confined to LENGTH [$F_{\mathrm{RATE}}(1,7)=95.69$, $p<0.001$] and EMOTION [$F_{\mathrm{RATE}}(3,21)=28.14$, $p<0.001$]. Long utterances were produced at a faster rate than short utterances. The EMOTION main effect [depicted in Fig. 6(a)] was ex-

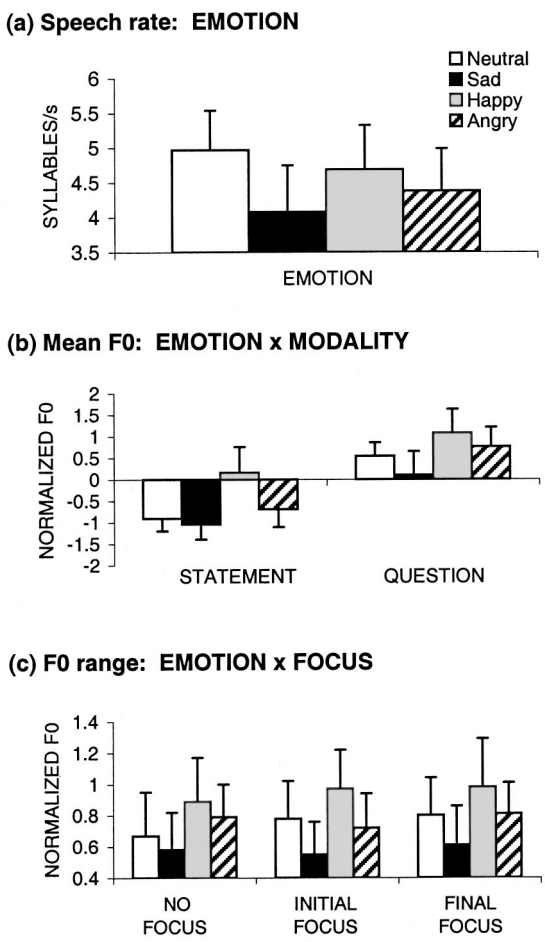

FIG. 6. Influence of emotion on three ''global'' measures of prosody as a function of relevant linguistic properties of the stimuli (averaged across eight speakers).

plained by an accelerated rate of delivery for neutral and happy utterances (which did not differ) relative to angry and sad utterances (which also did not differ). No interactions emerged from this analysis.

Analysis of mean utterance $f0$ yielded three related interactions centred around the MODALITY factor: LENGTH×MODALITY [$F_{\mathrm{MEAN}f0}(1,7)=64.67$, $p<0.001$]; FOCUS×MODALITY [$F_{\mathrm{MEAN}f0}(2,14)=28.03$, $p<0.001$]; and EMOTION×MODALITY [$F_{\mathrm{MEAN}f0}(3,21)=6.73$, $p<0.01$]. *Post hoc* comparisons indicated that in all conditions of each of these three interactions, questions were associated with a higher overall $f0$ than statements. However, for LENGTH×MODALITY, short utterances displayed a higher $f0$ mean than long utterances when formulated as a question but not when produced as a statement. The FOCUS×MODALITY interaction revolved around stimuli produced with *sentence-initial* focus: these tokens were characterized by a lower (statements) or higher (questions) mean $f0$ than matched stimuli in most other focus conditions. These trends were undoubtedly linked to the large post-focus fall or rise following FOC1 in this focus context, described earlier. The relationship between emotion and modality is presented in Fig. 6(b). It is shown that happy sentences exhibited a consistently higher mean $f0$ than sad or neutral sentences in both statements and questions. For statements, happy sentences were also higher than angry sentences, which did not differ from neutral or sad. For questions, angry and neutral sentences were associated with a comparable mean $f0$ and were both significantly higher than sad sentences.

Finally, analysis of $f0$ range yielded effects for EMOTION [$F_{f0\mathrm{RANGE}}(3,21)=20.27$, $p<0.001$] and EMOTION×FOCUS [$F_{f0\mathrm{RANGE}}(6,42)=4.88$, $p=0.01$], the interaction being supplied by Fig. 6(c). Happy utterances contained greater $f0$ variation than all other emotional modes examined (angry, neutral, sad). Angry utterances also demonstrated a larger range than sad utterances, and neutral utterances exceeded sad utterances in certain focus contexts (initial- and final-focus). Considered another way, it appeared that utterances with overt emotional features (happy, sad, angry) were associated with a relatively uniform $f0$ range across the three levels of focus, whereas the $f0$ range of neutral utterances was more dependent on focus context, displaying a larger range in contexts where focus was realized in the utterance (initial- and final-focus) than where it was not (no focus).

IV. DISCUSSION

Prosodic aspects of discourse provide a vehicle for speakers to concurrently express linguistic and emotional attributes of their mental states to listeners. Despite longstanding recognition of the multiple intentions represented in suprasegmental structure (e.g., Bolinger, 1986; Ladd, 1996), the acoustic realization of these conjoint intentions in spoken utterances has been accorded little empirical attention (cf. Cosmides, 1983; McRoberts *et al.*, 1995).

Discussion initially revolves around linguistic properties of the utterances sampled. There was a robust tendency for focused words in matched statements and questions to display higher $f0$ peaks and extended vowel segments than corresponding words produced without focus, supporting well-established data (Cooper *et al.*, 1985; Eady and Cooper, 1986; Ferreira, 1993; Weismer and Ingrisano, 1979). As was also expected, the local distribution of cues marking focus at designated keyword sites (FOC1 and FOC2 in the present study) was influenced by the *position* of focus realization in the utterance (Eady and Cooper, 1986; Weismer and Ingrisano, 1979). For both duration and $f0$, acoustic differentiation of the initial word of utterances (FOC1) was contingent on whether FOC1 was focused or not and did not vary for either parameter in the ''final'' and ''no'' focus speaking contexts. This description contrasts with how duration and $f0$ were regulated in sentence final position (FOC2); here, encoders differentiated ''focused'' elements from those in both ''unfocused'' contexts in the predicted manner. However, they also reduced the temporal and $f0$ prominence of FOC2 when occurring ''post-focus'' (i.e., in the ''initial'' focus context) relative to the cues provided in utterances that did not contain focus (Eady and Cooper, 1986; Weismer and Ingrisano, 1979 for duration only). Patterns observed for sentence-final words imply the operation of a ''retroactive'' strategy governing a speaker's on-line decision about focus production; the semantic value of initial focused words is reinforced by diminishing the duration and $f0$ salience of

later, potentially contrastive forms (Cooper *et al.*, 1985; Eady *et al.*, 1986). However, the same encoders did not supply ''anticipatory'' cues to focus produced at the end of utterances by attenuating the acoustic prominence of concepts at the onset of the stimulus (Eady and Cooper, 1986; Weismer and Ingrisano, 1979).

Further consideration of the keyword measures argues that for duration, focus effects are not always strictly localized to the target word as described for previous stimuli (Cooper *et al.*, 1985; Eady and Cooper, 1986). Rather, such effects manifest more broadly throughout carrier utterances as a function of stimulus characteristics. As recognized by Eady *et al.* (1986), acoustic studies that have concentrated their analyses on relatively long (10–12 syllables) utterances have reported localized durational effects for focus (Cooper *et al.*, 1985; Eady and Cooper, 1986) whereas those examining shorter (5–7 syllables) stimuli have not (Folkins *et al.*, 1975; Weismer and Ingrisano, 1979; also trends in Eady *et al.*, 1986). The impact of sentence length was directly testable in the current study which elicited six-syllable (short) and ten-syllable (long) utterances from a uniform sample of encoders.

Results confirmed that utterance length contributes to the temporal properties of focused versus unfocused words, although not entirely in the projected manner. ''Long'' utterances resembling those employed by Eady and Cooper (1986) and others exhibited fewer instances where the duration of neighboring words was modified due to focus placement (especially for final keywords). This pattern is suggestive of localized cue use. However, post-focal reduction in the duration of FOC2, a nonlocalized effect, was observed for both long and short utterances. Moreover, content words directly preceding FOC2 in long stimuli (*teapot* in *Mary sold the teapot for a dollar*) were sensitive to focus position, displaying reduced durations in *pre*-focal environments (Cooper *et al.*, 1985). Thus neither short nor long utterances contained purely localized alterations in duration due to the operation of linguistic focus. Possibly, syntactic-semantic attributes of experimental stimuli in which such pragmatic decisions are marked—which vary considerably across studies in this literature (Cooper *et al.*, 1985; Eady and Cooper, 1986; Folkins *et al.*, 1975; Weismer and Ingrisano, 1979)—further dictate the extent to which speakers manipulate prosodic aspects of the message to communicate focus, influencing the localizability of associated cues (particularly for long stimuli).

The $f0$ correlates of focus were not expected to be localized in the utterance but have a pervasive impact on the shape of the intonation contour. This assumption was borne out for both statements and questions, particularly in the initial-focus condition where marked falling (statement) or rising (question) $f0$ excursions characterized sentence-initial words and their immediate post-focus environment (Cooper *et al.*, 1985; Eady and Cooper, 1986; O'Shaughnessy, 1979). These abrupt transitions influenced the $f0$ of all subsequent points of the contour topline for this one focus condition when matched statements and questions in the other focus conditions were compared. In neutral- and final-focus utterances, the divergence between falling (statement) and rising (question) contours progressed more gradually and did not always differ significantly until the ''terminal'' point (i.e., final 150 ms) where all speakers reliably distinguished these two speech acts (Eady and Cooper, 1986; Hadding-Koch and Studdert-Kennedy, 1964; Lieberman, 1967; O'Shaugnessy, 1979; Studdert-Kennedy and Hadding, 1973).

Thus discontinuities in the $f0$ topline may furnish powerful clues to focus position within statements and questions. This being said, it is interesting to consider that for interrogative utterances, speakers did not furnish *any* statistically detectable $f0$ cues that biased the position of focus in the utterance when measures were limited to sentence-initial or sentence-final keywords (Eady and Cooper, 1986). A potentially related trend was noted for keyword duration, where acoustic differences among the three focus conditions were smaller and fewer in number for questions than for statements (although recall that focus lengthening was distinctive in both modality types). These findings imply that English utterances produced with the terminal yes/no interrogation rise place observable constraints on a speaker's capacity to modulate $f0$ and duration to realize focus using the range of cues typical of matched words produced in declarative utterances (McRoberts *et al.*, 1995). Presumably, such prosodic differences have a negligible effect on *listeners'* ability to infer what linguistic elements are highlighted by the speaker in declarative versus interrogative speech acts, although the present study was not designed to address this issue (perceptual evaluation of speakers was restricted to declarative renditions of the stimuli).

Specifying the acoustic form of 'basic' emotional displays was estimated via three frequently cited parameters of emotional vocalizations: mean $f0$, $f0$ range, and speech rate (Frick, 1985; Scherer, 1986). In broad terms, results obtained here for elderly speakers accord with past descriptions of younger subjects engaged in simulated or naturalistic displays of vocal emotion. Specifically, sad utterances were spoken with a relatively slow rate, a low mean $f0$ and highly restricted $f0$ range (Breitenstein *et al.*, in press; Huttar, 1968; Johnson *et al.*, 1986; Williams and Stevens, 1972). Neutral utterances were reliably faster than sad utterances (Breitenstein *et al.*, in press; Fairbanks and Hoaglin, 1941; Williams and Stevens, 1972) with a somewhat higher mean $f0$ and extended $f0$ range than sad in many conditions. Happy utterances exhibited a higher mean $f0$ and broader $f0$ range than sad or neutral stimuli (Fonagy, 1978; Huttar, 1968; Scherer, 1974). In the current data set, angry utterances also displayed a higher overall $f0$ and broader $f0$ range than sad and neutral stimuli, and happy exceeded angry stimuli on these dimensions in many speaking conditions (cf. Williams and Stevens, 1972, for a comparison of happy and angry).[3] Finally, happy vocalizations tended to be faster than angry or sad utterances for the present stimuli (happy=neutral), although attempts to characterize the rate of ''happy'' voices have led to mixed conclusions overall (Murray and Arnott, 1993). As an overview, the current set of measures—while clearly inadequate descriptors of the array of vocal features contributing to emotional communication for any speaker group (e.g., ''angry'' voices may rely heavily on aspects of voice quality; e.g., Cummings and

Clements, 1995; Fonagy and Magdics, 1963; Johnson *et al.*, 1986)—were nonetheless relatively successful in separating the four emotional targets at the acoustic level, and in a manner which was largely predicted by past research (Murray and Arnott, 1993; Scherer, 1986).

The impact of simulated emotion on the acoustic correlates of linguistic focus and utterance modality was of central interest here. First, note that emotion was not expected to substantially alter the pattern of acoustic associations between focused and unfocused words beyond that related to utterance modality and this prediction was confirmed. However, the especially vital contribution of pitch/$f0$ in communicating both linguistic and emotional aspects of target messages yielded meaningful interactions among levels of emotion, focus, and modality when $f0$ parameters of sentence-initial and sentence-final keywords were inspected. Notably, speaking conditions in which encoders were required to modulate $f0$ to convey both focus and the ''marked'' terminal rise of questions in English (Lieberman, 1967) were associated with an *elimination* of $f0$ distinctions conducive to different emotional interpretations at the location of key content words. Such tendencies were not witnessed for declarative speech acts which seemed to invest speakers with greater flexibility in manipulating emotion-related $f0$ parameters in most focus contexts, permitting the relatively distinctive array of acoustic gestures typical of these four emotions (Murray and Arnott, 1993). Presumably, past data on the vocal characterization of emotion have been derived almost exclusively for declarative utterances, explaining previous failures to detect such modality-related differences.

Thus there are tentative indications in the data that modulating $f0$ parameters that preserve the communicative features of both focus position and yes/no questions mitigate the scope of *emotion*-related $f0$ differences permitted on key content words when compared to those expressed in declarative utterances. Certainly, this is not to claim that emotion markers were absent in interrogative contours; in addition to speech rate and other unexplored parameters of emotion that were undoubtedly manifest in both types of speech acts, analysis of mean utterance $f0$ revealed reliable cross-emotion trends that were qualitatively similar for statements and questions (e.g., happy exhibited a consistently higher overall $f0$ than sad and neutral). As such, it may be concluded that emotion-related $f0$ characteristics of declarative and interrogative speech acts are encoded consistently as a global function of the intonation contour (e.g., Ladd *et al.*, 1985; Scherer, 1986), but that for interrogative utterances, emotional differences in $f0$ are comparatively opaque at points where contrastive focus is operational. At such points in the contour, $f0$ features of the four emotions tend to converge.

Given these interactive effects on $f0$, one may speculate that the need to assign contrastive focus within rising intonation contours—joint linguistic intentions achieved through regulation of vocal fold tension due to laryngeal (primarily cricothyroid) activity (Atkinson, 1973; Lieberman *et al.*, 1970)—places unique constraints on English speakers in the modulation of $f0$ for *additional* purposes such as emotional inflection, at least in simulated speaking contexts. This explanation, which appeals to the idea of the ''prosodic load'' of an utterance and its impact on speakers, resembles an account proposed by McRoberts and colleagues (1995) in a related study of simple English utterances (e.g., *November*). Those authors attributed a trade-off in the ability to program $f0$ for simultaneous *linguistic* purposes (stress and interrogation rise) to articulatory constraints imposed on their four male speakers. However, requiring the same speakers to manipulate the affective (positive/negative) valence of interrogative stimuli did not yield a similar $f0$ trade-off for their stimuli, suggesting to the investigators that affective and linguistic programming of $f0$ may be functionally separate (McRoberts *et al.*, 1995).

Many important differences characterize the methods and stimuli employed here and in this earlier investigation, especially the form in which emotionally inflected speech was characterized and evaluated for experimental stimuli. Furthermore, McRoberts and colleagues did not require their subjects to manipulate focus, speech acts, *and* emotional attributes in tandem (the effects of focus and affect on interrogative contours were evaluated in separate conditions). As such, the prosodic load and its presumed effect on $f0$ implementation was greater for the current set of speakers. Indeed, it is precisely this need to combine ''marked'' linguistic-prosodic forms and specific emotional targets that imposed the most severe articulatory limitations on the present speakers, implying that linguistic and emotional uses of $f0$ are not always functionally separate in speech production. Further work on the interaction between emotional and linguistic aspects of prosody will help shed light on these promising, yet preliminary hypotheses.

Brief commentary is reserved for data on the form of ''happy'' utterances. In this condition, the duration of sentence-final keywords was robustly longer than in the other three emotional modes independent of a broad range of variables (focus position, modality, or sentence length). Moreover, the $f0$ glide of questions terminated at a significantly higher point when speakers were happy than in some other emotional modes (in contrast, the endpoint of statements was a stable feature across emotional contexts; Liberman and Pierrehumbert, 1984). Differences in the terminal point of the interrogative rise, although consistent with described trends in the $f0$ data for happy (and formulations proposed by Ohala, 1983), are again curious in light of McRoberts *et al.*'s (1995) data; those authors reported that the positive or negative valence of short utterances had no effect on the extent of $f0$ rise marking yes/no questions. This discrepancy will again benefit from future testing employing a relatively broad range of stimuli.

Evidence of the impact of emotion on aspects of intonation and word duration underscore the diverse ways in which humans modulate a small set of acoustic parameters to communicate complex linguistic and emotional intentions. By the same token, findings are limited by the restricted purview of the acoustic measures employed here which do not capture the range of cues available to speakers to signal such intentions (particularly for emotional expression; e.g., Scherer, 1986). Despite relative success in simulating three

prototypical emotional contexts for analysis, future endeavors which address the influence of emotion on intonation will benefit from recordings of naturalistic emotional displays, and a broader range of emotions to weigh against the current data (e.g., ''fear'' and ''disgust'' may constitute basic human emotions; Ekman, 1992; Izard, 1977). The important role of gender on the ability to encode emotion is also showcased here, as the four individuals who were best able to communicate emotional qualities to listeners were all female (Sobin and Alpert, 1999; Zuckerman *et al.*, 1975).

ACKNOWLEDGMENTS

Special thanks to Marta Fundamenski, Nazma Mohammed, and Anita Shuper for help in subject testing and manuscript preparation, and to Dr. A. Löfqvist and two anonymous reviewers for valuable comments received on an earlier version of this paper. This research was supported by the Québec Fonds pour la formation des chercheurs et l'aide à la recherche (FCAR) and a Fraser, Monat, and McPherson scholarship awarded by the McGill Faculty of Medicine.

[1]It is assumed that speech samples elicited by Cooper, Eady, and others resemble emotionally ''neutral'' speech generated in the current paradigm.

[2]Overall values are not inconsistent with related studies obtaining perceptual ratings of prosody from untrained listeners using a forced-choice paradigm (e.g., Williams and Stevens, 1972).

[3]Frick (1985) and others have distinguished between anger which represents ''frustration'' versus ''threat,'' with only the former correlating with raised pitch (Fairbanks and Pronovost, 1939; Williams and Stevens, 1972). Contextual cues provided in this study biased responses indicative of ''frustrated'' anger, yielding the predicted rise in mean *f*0.

Atkinson, J. (**1973**). ''Aspects of intonation in speech: implications from an experimental study of fundamental frequency,'' unpublished doctoral dissertation, University of Connecticut.

Bachorowski, J. (**1999**). ''Vocal expression and perception of emotion,'' Curr. Dir. Psychol. Sci. **8**, 53–57.

Bachorowski, J., and Owren, M. J. (**1995**). ''Vocal expression of emotion: Acoustic properties of speech are associated with emotional intensity and context,'' Psycholog. Sci. **6**, 219–224.

Banse, R., and Scherer, K. (**1996**). ''Acoustic profiles in vocal emotion expression,'' J. Personality Soc. Psychol. **70**, 614–636.

Behrens, S. J. (**1988**). ''The role of the right hemisphere in the production of linguistic stress,'' Brain Lang. **33**, 104–127.

Bolinger, D. (**1978**). ''Intonation across languages,'' in *Universals of Human Language*, edited by J. H. Greenberg (Stanford University Press, Stanford, CA), pp. 471–524.

Bolinger, D. (**1986**). *Intonation and its Parts* (Stanford University Press, Stanford, CA).

Bolinger, D. L. (**1955**). ''Intersections of stress and intonation,'' Word **11**, 195–203.

Bolinger, D. L. (**1958**). ''A theory of pitch accent in English,'' Word **14**, 109–149.

Breitenstein, C., Van Lancker, D., and Daum, I. (in press). ''The contribution of speech rate and pitch variation to the perception of vocal emotions in a German and an American sample,'' Cognition and Emotion.

Brown, W. S., and McGlone, R. E. (**1974**). ''Aerodynamic and acoustic study of stress in sentence productions,'' J. Acoust. Soc. Am. **56**, 971–974.

Cooper, W., and Sorensen, J. (**1981**). *Fundamental Frequency in Sentence Production* (Springer, New York).

Cooper, W. E., Eady, S. J., and Mueller, P. R. (**1985**). ''Acoustical aspects of contrastive stress in question-answer contexts,'' J. Acoust. Soc. Am. **77**, 2142–2156.

Cosmides, L. (**1983**). ''Invariances in the acoustic expression of emotion during speech,'' J. Exp. Psychol. **9**, 864–881.

Cummings, K. E., and Clements, M. A. (**1995**). ''Analysis of the glottal excitation of emotionally styled and stressed speech,'' J. Acoust. Soc. Am. **98**, 88–98.

Davitz, J. R. (**1964**). ''Auditory correlates of vocal expressions of emotional meanings,'' in *The Communication of Emotional Meaning*, edited by J. R. Davitz (McGraw-Hill, New York), pp. 101–112.

Denes, P. (**1959**). ''A preliminary investigation of certain aspects of intonation,'' Lang. Speech **2**, 107–122.

Eady, S. J., and Cooper, W. E. (**1986**). ''Speech intonation and focus location in matched statements and questions,'' J. Acoust. Soc. Am. **80**, 402–415.

Eady, S. J., Cooper, W. E., Klouda, G. V., Mueller, P. R., and Lotts, D. W. (**1986**). ''Acoustical characteristics of sentential focus: Narrow vs broad and single vs. dual focus environments,'' Lang Speech **29**, 233–251.

Eefting, W. (**1990**). ''The effect of ''information value'' and ''accentuation'' on the duration of Dutch words, syllables, and segments,'' J. Acoust. Soc. Am. **89**, 412–424.

Ekman, P. (**1992**). ''An argument for basic emotions,'' Cognition and Emotion **6**, 169–200.

Ekman, P., Levenson, R. W., and Friesen, W. V. (**1983**). ''Autonomic nervous system activity distinguishes among emotions,'' Science **221**, 1208–1210.

Fairbanks, G., and Hoaglin, L. W. (**1941**). ''An experimental study of the durational characteristics of the voice during the expression of emotion,'' Speech Monographs **8**, 85–90.

Fairbanks, G., and Pronovost, W. (**1939**). ''An experimental study of the pitch characteristics of the voice during the expression of emotion,'' Speech Monographs **6**, 87–104.

Ferreira, F. (**1993**). ''Creation of prosody during sentence production,'' Psychol. Rev. **100**, 233–253.

Folkins, J. W., Miller, C. J., and Minifie, F. D. (**1975**). ''Rhythm and syllable timing in phrase level stress patterning,'' J. Speech Hear. Res. **18**, 739–753.

Fonagy, I. (**1978**). ''A new method of investigating the perception of prosodic features,'' Lang Speech **21**, 34–49.

Fonagy, I., and Magdics, K. (**1963**). ''Emotional patterns in intonation and music,'' Z. Phonetik **16**, 293–326.

Fowler, C., and Housum, J. (**1987**). ''Talkers' signaling of ''new'' and ''old'' words in speech and listeners' perception and use of the distinction,'' J. Memory Lang. **26**, 489–504.

Frick, R. W. (**1985**). ''Communicating emotion: The role of prosodic features,'' Psychol. Bull. **97**, 412–429.

Fry, D. B. (**1955**). ''Duration and intensity as physical correlates of linguistic stress,'' J. Acoust. Soc. Am. **27**, 765–768.

Fry, D. B. (**1958**). ''Experiments in the perception of stress,'' Lang Speech **1**, 126–152.

Gandour, J., Larsen, J., Dechongkit, S., Ponglorpisit, S., and Khunadorn, F. (**1995**). ''Speech prosody in affective contexts in Thai patients with right hemisphere lesions,'' Brain Lang. **51**, 422–443.

Hadding-Koch, K., and Studdert-Kennedy, M. (**1964**). ''An experimental study of some intonation contours,'' Phonetica **11**, 175–185.

Huttar, G. (**1968**). ''Relations between prosodic variables and emotions in normal American English utterances,'' J. Speech Hear. Res. **11**, 467–480.

Izard, C. E. (**1977**). *Human Emotions* (Plenum, New York).

Johnson, W. F., Emde, R. N., Scherer, K. R., and Klinnert, M. D. (**1986**). ''Recognition of emotion from vocal cues,'' Arch. Gen. Psychiatry **43**, 280–283.

Klatt, D. H. (**1976**). ''Linguistic uses of segmental duration in English: Acoustic and perceptual evidence,'' J. Acoust. Soc. Am. **59**, 1208–1221.

Ladd, D. R. (**1996**). *Intonational Phonology* (Cambridge University Press, Cambridge).

Ladd, D. R., Silverman, K. E. A., Tolkmitt, F., Bergmann, G., and Scherer, K. R. (**1985**). ''Evidence for the independent function of intonation contour type, voice quality, and *F*0 range in signaling speaker effect,'' J. Acoust. Soc. Am. **78**, 435–444.

Lea, W. (**1977**). ''Acoustic correlates of stress and juncture,'' in *Studies in Stress and Accent*, edited by L. M. Hyman (Dept. of Linguistics, U.C.L.A., Los Angeles), pp. 83–120.

Liberman, M., and Pierrehumbert, J. (**1984**). ''Intonational invariance under changes in pitch range and length,'' in *Language Sound and Structure*, edited by M. Aronoff and R. Oehrle (MIT Press, Cambridge, MA), pp. 157–233.

Lieberman, P. (**1967**). *Intonation, Perception, and Language* (MIT Press, Cambridge, MA).

Lieberman, P., and Michaels, S. B. (**1962**). ''Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech,'' J. Acoust. Soc. Am. **34**, 922–927.

Lieberman, P., Sawashima, M., Harris, K., and Gay, T. (**1970**). ''The articulatory implementation of the breath-group and prominence: Cricothyroid muscular activity in intonation,'' Language **46**, 312–327.

Majewski, W., and Blasdell, R. (**1968**). ''Influence of fundamental frequency cues on the perception of some synthetic intonation contours,'' J. Acoust. Soc. Am. **45**, 450–457.

Max, L., and Onghena, P. (**1999**). ''Some issues in the statistical analysis of completely randomized and repeated measures designs for speech, language, and hearing research,'' J. Speech Lang Hear Res. **42**, 261–270.

McClean, M. D., and Tiffany, W. R. (**1973**). ''The acoustic parameters of stress in relation to syllable position, speech loudness and rate,'' Lang. Speech **16**, 283–290.

McRoberts, G. W., Studdert-Kennedy, M., and Shankweiler, D. P. (**1995**). ''The role of fundamental frequency in signaling linguistic stress and affect: Evidence for a dissociation,'' Percept. Psychophys. **57**, 159–174.

Mertus, J. (**1989**). *BLISS User's Manual* (Brown University, Providence, RI).

Morton, J., and Jassem, W. (**1965**). ''Acoustic correlates of stress,'' Lang. Speech **8**, 159–181.

Murray, I. R., and Arnott, J. L. (**1993**). ''Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion,'' J. Acoust. Soc. Am. **93**, 1097–1108.

Ohala, J. (**1983**). ''Cross-language use of pitch: An ethological view,'' Phonetica **40**, 1–18.

O'Shaughnessy, D. (**1979**). ''Linguistic features in fundamental frequency patterns,'' J. Phonetics **7**, 119–145.

Ouellette, G. P., and Baum, S. R. (**1993**). ''Acoustic analysis of prosodic cues in left- and right-hemisphere-damaged patients,'' Aphasiology **8**, 257–283.

Pakosz, M. (**1983**). ''Attitudinal judgments in intonation: Some evidence for a theory,'' J. Psycholinguist. Res. **12**, 311–326.

Pell, M. D. (**1999a**). ''Fundamental frequency encoding of linguistic and emotional prosody by right hemisphere-damaged speakers,'' Brain Lang. **69**, 161–192.

Pell, M. D. (**1999b**). ''Some acoustic correlates of perceptually ''flat affect'' in right-hemisphere damaged speakers,'' Brain Cogn. **40**, 219–223.

Pell, M. D. (**1999c**). ''The temporal organization of affective and nonaffective speech in patients with right-hemisphere infarcts,'' Cortex **35**, 455–477.

Rose, P. (**1987**). ''Considerations in the normalization of the fundamental frequency of linguistic tone,'' Speech Commun. **6**, 343–351.

Ross, E. D., Edmondson, J. A., and Seibert, G. B. (**1986**). ''The effect of affect on various acoustic measures of prosody in tone and nontone languages: A comparison based on computer analysis of voice,'' J. Phonetics **14**, 283–302.

Ryalls, J., Le Dorze, G., Lever, N., Ouellet, L., and Larfeuil, C. (**1994**). ''The effects of age and sex on speech intonation and duration for matched statements and questions in French,'' J. Acoust. Soc. Am. **95**, 2274–2276.

Scherer, K. R. (**1974**). ''Acoustic concomitants of emotional dimensions: Judging affect from synthesized tone sequences,'' in *Non-Verbal Communication*, edited by S. Weitz (Oxford University Press, New York), pp. 105–111.

Scherer, K. R. (**1986**). ''Vocal affect expression: A review and a model for future research,'' Psychol. Bull. **99**, 143–165.

Siegman, A. W., and Boyle, S. (**1993**). ''Voices of fear and anxiety and sadness and depression: The effects of speech rate and loudness on fear and anxiety and sadness and depression,'' J. Abnorm. Psychol. **102**, 430–437.

Siegman, A. W., Dembroski, T. M., and Crump, D. (**1992**). ''Speech rate, loudness, and cardiovascular reactivity,'' J. Behav. Med. **15**, 519–532.

Sobin, C., and Alpert, M. (**1999**). ''Emotion in speech: The acoustic attributes of fear, anger, sadness, and joy,'' J. Psycholinguist. Res. **23**, 347–365.

Streeter, L., MacDonald, N., Apple, W., Krause, R., and Galotti, K. (**1983**). ''Acoustic and emotional indicators of emotional stress,'' J. Acoust. Soc. Am. **73**, 1354–1360.

Studdert-Kennedy, M., and Hadding, K. (**1973**). ''Auditory and linguistic processes in the perception of intonation contours,'' Lang. Speech **16**, 293–313.

Turk, A. E., and Sawusch, J. R. (**1996**). ''The processing of duration and intensity cues to prominence,'' J. Acoust. Soc. Am. **99**, 3782–3790.

Uldall, E. (**1960**). ''Attitudinal meanings conveyed by intonation contours,'' Lang. Speech **3**, 223–234.

Weismer, G., and Ingrisano, D. (**1979**). ''Phrase-level timing patterns in English: Effects of emphatic stress location and speaking rate,'' J. Speech Hear. Res. **22**, 516–533.

Williams, C. E., and Stevens, K. N. (**1972**). ''Emotions and speech: Some acoustical correlates,'' J. Acoust. Soc. Am. **52**, 1238–1250.

Zuckerman, M., Lipets, M., Hall, J., and Rosenthal, R. (**1975**). ''Encoding and decoding nonverbal cues of emotion,'' J. Pers. Soc. Psychol. **32**, 1068–1076.