Enhanced contrast for vowels in utterance focus: A cross-language study

Jessica F. Hay, Momoko Sato, Amy E. Coren, Cheryl L. Moran, and Randy L. Diehl
Department of Psychology and Center for Perceptual Systems, University of Texas at Austin, 1 University Station A8000, Austin, Texas 78712-0187

(Received 3 May 2005; revised 10 February 2006; accepted 10 February 2006)

The present study examined several potential distinctiveness-enhancing correlates of vowels produced in utterance focus by talkers of American English, French, and German. These correlates included possible increases in vowel space size, in formant movement within individual vowels, and in duration variance among vowels. Each language group enhanced the distinctiveness of vowels in [+focus] context but used somewhat differing means to achieve this. All three groups used spectral differences, but only German talkers used durational differences, to enhance distinctiveness. The results suggest that the amount of distinctiveness enhancement of a vowel property in [+focus] context is positively related to the between-category variation of that property in [−focus] context. Thus, consistent with the theory of adaptive dispersion, utterance clarity appears to vary directly with information content. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2184226]

PACS number(s): 43.70.Bk, 43.70.Fq, 43.70.Kv [ARB] Pages: 3022–3033

I. INTRODUCTION

In this study, we examined several possible distinctiveness-enhancing correlates of focally prominent syllables in American English, French, and German. The aim was to assess cross-language commonalities and differences in how phonological distinctiveness of vowels is increased in [+focus] contexts.

Jakobson *et al.* (1963) famously remarked that "we speak to be heard in order to be understood" (p. 13). In his theory of adaptive dispersion, Lindblom (1986, 1990) elaborated this point as follows: the aim of talkers is to produce utterances that are intelligible to listeners, but to do so with as little effort as necessary. In other words, talkers try to achieve sufficient, rather than maximal, distinctiveness in articulating vowels and consonants, and they thus tend to vary their utterances from reduced (hypospeech) forms to clear (hyperspeech) forms according to the communication conditions that apply.

A variety of findings are consistent with Lindblom's account. Communication conditions that have been shown to influence how clearly talkers articulate speech sounds include factors pertaining to the communication channel, the listener, and the information content of the message.

Communication channel and listener factors: An early study by Lombard (1911) demonstrated that talkers produce clearer and louder speech in the presence of background noise, a result that has since been replicated and extended (Lane *et al.*, 1970; Lane and Tranel, 1971). As the distance to the listener increases, talkers increase their vocal effort, although not enough to compensate fully for the distance effect on signal intensity (Traunmüller and Eriksson, 2000). In conversational speech, talkers tend to speak more clearly when the line of sight between talker and listener is obstructed, and the listener is thus deprived of visual speech reading cues, than when the talker and listener are able to see each other (Anderson *et al.*, 1997). Picheny *et al.* (1985, 1986) instructed talkers to produce either ordinary conversational speech or clear speech as if "trying to communicate in a noisy environment or to an impaired listener" (Picheny *et al.*, 1985, p. 97). The clear speech was produced at a slower rate, a higher intensity, and with less vowel and consonant reduction than the conversational speech. As expected, the clear speech was more intelligible to hearing impaired listeners than the conversational speech.[1] Talkers tend to produce more widely dispersed vowel categories, yielding less acoustic overlap among categories, when speaking to non-native speaking listeners than when speaking to native speaking listeners (Knoll and Uther, 2004). Such results suggest that talkers modulate their speech clarity according to demands imposed by the communication channel and by the listener.

Information content of the message: Perhaps the strongest evidence that talkers aim for sufficient—rather than maximal—distinctiveness is pervasive findings that utterance clarity tends to vary positively with information content: (a) Words produced in isolation are more accurately identified than the same words excised from running speech (Pollack and Pickett, 1963). The excised word tokens are more reduced, presumably because of their greater predictability (i.e., lower information content) in the original contexts. (b) Words excised from stereotyped phrases or adages (e.g., "A stitch in time saves [*nine*]") are less intelligible than the same words excised from less redundant contexts (e.g., "The number that you will hear is [*nine*]") (Lieberman, 1963; see Hunnicutt, 1985, for a partial replication and qualification of this result). (c) Words produced for the first time in the course of a monolog are more intelligible when excised from their contexts than words produced for the second time (Fowler and Housum, 1987). (d) Vowels are more clearly articulated in words that occupy dense lexical neighborhoods and that are low in frequency of occurrence relative to their nearest phonological neighbors (Wright, 2003). Higher neighborhood density and lower frequency of occurrence yield higher information content, and both are associated with lower word recognition accuracy (Luce, 1986). (e) Vowels are more reduced in closed class function words (e.g., the auxiliary verb "can") than in open class content words (e.g., the noun "candy") (van Bergem, 1993). (f) Vowels in lexically unstressed syllables are more reduced (i.e., shorter in duration and less distant from schwa) than vowels in lexically stressed syllables (van Bergem, 1993). Cross-linguistically, vowels in stressed syllables tend to carry more information because of the greater variety of vowels permitted in stressed contexts (Altman and Carter, 1989). (g) Consonants at the beginning of prosodic domains such as words or phrases are more clearly articulated and less confusable than later-occurring consonants, which are generally more predictable (Jun, 1993; Byrd, 1994; Browman and Goldstein, 1995; Redford and Diehl, 1999; Keating *et al.*, 1999; Cho and Keating, 2001). (h) Vowels in [+focus] context (i.e., in portions of an utterance that contain *new*, as opposed to *given*, information) are produced with greater separation between vowel categories than those in [−focus] context (van Bergem, 1993). (i) In languages with vowel length distinctions, duration differences between short and long vowels tend to be enhanced in [+focus] context. Such enhancement has been observed in Swedish (Heldner and Strangert, 2001), Arabic (de Jong and Zawaydeh, 2002), and Serbian (Smiljanic, 2004). (j) Phonetic correlates of [voice] contrasts show enhanced acoustic distinctiveness in [+focus] context (Cho and McQueen, 2005).

In examples (a)–(j), information content was varied and/or assessed qualitatively, that is, in the form of ordinal comparisons. Taking a more quantitative approach, van Son and Pols (2003) used a large sample of spoken Dutch to estimate the information (in bits) carried by individual phonemes in word and sentence context. Across several speaking styles, they found positive correlations between information content and two measures of speech clarity—duration and spectral distinctiveness. Although these correlations were highly significant statistically, the R values were fairly small (<0.3), owing perhaps to the many other sources of durational and spectral variance in natural productions.

As this last study suggests, talker compensation for varying communication conditions is far from perfect. In some cases (e.g., varying background noise, varying distance separating talker and listener, and varying degrees of hearing loss), there may be physical or perceptual limits to maintaining uniform levels of intelligibility. Moreover, in cases of varying information content, talkers may lack the cognitive means to estimate accurately the information available to listeners on a segment-by-segment basis, and thus articulatory compensation may be rather crude. For example, Billerey-Mosier (2000) reported no systematic difference in vowel duration as a function of whether the vowel occurred at the carrier word's uniqueness point (i.e., the earliest point at which the word differs from all lexical neighbors sharing the same initial sequence of phonemes) or after the uniqueness point, where the vowel carries no information. Also, in the study by Anderson *et al.* (1997) cited earlier, the reduction in talker clarity that occurred with an unobstructed line of sight between talker and listener was not closely linked to the moment-to-moment looking behavior of the listener. Finally, talkers may sometimes have difficulty distinguishing what is given information for themselves and what is given information for the listener. For example, Bard *et al.* (2000) reported that when talkers named map landmarks in spontaneous speech to two successive listeners, the key words were produced less clearly on repetition even though the second listener had not heard the original version. Despite such findings, however, it is evident that Lindblom's notion of sufficient contrast accounts at least qualitatively for an impressive range of data.

The present study examines how production of vowels is affected by whether or not they occur in utterance focus, a variable associated with differences in information content. Declarative sentences can typically be analyzed into two kinds of grammatical constituents: those carrying information already known to the listener (given information) and those carrying information not previously known to the listener (new information). The feature [+focus] applies to the latter kind of constituent, whereas [−focus] applies to the former.[2] In general, [+focus] corresponds to the constituent occupying the location of a wh-phrase in the answer to a wh-question:

$$\begin{aligned}&\text{Q: Who stole the money?}\\&\text{A: }[\text{Bob}]_F\text{stole the money.}\end{aligned} \tag{1}$$

The grammatical constituent marked as [+focus] usually includes the syllable that is intonationally most prominent in the utterance. Linguists have offered varying accounts of how focal prominence is assigned within the focused constituent (Chomsky and Halle, 1968; Ladd, 1980, 1996; Gussenhoven, 1983; Selkirk, 1984; Baart, 1987; Bolinger 1989). Typically, it is assumed to be assigned to the primary stressed syllable of the head (i.e., most important lexical item) of the focused constituent.

Phonetic investigations of various languages have described several correlates of focal prominence. Relative to stressed syllables in [−focus] context, stressed syllables in [+focus] context tend to be longer (Nooteboom, 1972; Cooper *et al.*, 1985; Eefting, 1991; Sluijter and van Heuven, 1995; Turk and Sawusch, 1997; Strangert and Heldner, 1998; Heldner and Strangert, 2001), to be more intense (Campbell, 1995; Sluijter and van Heuven, 1996a; Heldner *et al.*, 1999; Heldner, 2003), to have more high-frequency emphasis (Campbell, 1995; Sluijter and van Heuven, 1996b; Heldner *et al.*, 1999; Heldner, 2003), and to have a greater and more rapid change in fundamental frequency (O'Shaughnessy, 1979; Lyberg, 1979; 't Hart *et al.*, 1990). All of these correlates of focus are likely to contribute to intonational prominence and thus to draw the listener's attention to the most informative portion of an utterance (Ladd and Cutler, 1983).

Consistent with Lindblom's notion of sufficient contrast, other correlates of focal prominence may have a distinctiveness-enhancing role, in addition to any attention-directing function they serve. One such correlate is size of the vowel space: [+focus] syllables tend to show a greater separation of vowel categories within a formant frequency ($F1 \times F2$) space (van Bergem, 1993). It is also possible that [+focus] syllables are associated with greater formant movement within each vowel and/or greater duration variance across vowel categories. Several investigators (e.g., Nearey and Assmann, 1986; Hillenbrand *et al.*, 1995; Hillenbrand and Nearey, 1999) have reported that American English vowels tend to exhibit "vowel-inherent spectral change" (VISC), whereby formant frequencies change from a relatively steady-state region near the onset of the vowel to a different set of values near the offset of the vowel. As the label implies, the frequency changes are assumed to be characteristic of the vowel itself and not merely a by-product of coarticulation with initial and final consonants. Including VISC in the design of synthetic American English vowel stimuli significantly improves their identifiability by listeners (Hillenbrand and Nearey, 1999), and including it in the parametric description of natural American English vowels significantly improves categorization performance by statistical pattern recognition models (Nearey and Assmann, 1986). Similar pattern recognition results have been reported for monophthongs of Australian English (Watson and Harrington, 1999). Because VISC contributes to vowel identity in at least one language, it may be subject to enhancement in [+focus] context, yielding greater acoustic distinctiveness among vowels. Apart from spectral differences, vowel categories tend to vary in duration, and in many languages, vowel length is phonemic (Lehiste, 1970). It was noted earlier that such vowel length distinctions have been shown to be enhanced in [+focus] context for several languages (Heldner and Strangert, 2001; de Jong and Zawaydeh, 2002; Smiljanic, 2004).

In the present study, we examined vowel space size, amount of formant movement, and duration variance as possible distinctiveness-enhancing correlates of [+focus] syllables in American English, French, and German. The aim was to assess cross-language commonalities and differences in how phonological distinctiveness of vowels is increased in [+focus] context. The requirement of sufficient (rather than maximal) contrast allows for the possibility that different languages may exploit varying means of distinctiveness enhancement. Such variation may, of course, reflect differences across languages in the phonological contrasts used, but it may also reflect different weights assigned to the various correlates of a given contrast. For example, a feature correlate that carries more information in language A than in language B may be more heavily exploited for distinctiveness enhancement in language A.

II. METHOD

A. Talkers

American English talkers were 12 undergraduates (six male and six female) from the University of Texas at Austin, all of whom spoke with a Central Texas dialect. They ranged in age from 18 to 22 years old. French talkers were 12 undergraduates (six male and six female) from the Université René Descartes, Paris, France. All spoke with a standard Parisian dialect, and they ranged in age from 18 to 29 years old. German talkers were 14 adults (seven male and seven female) who were currently living in or visiting the central Texas area and who ranged in age from 15 to 38. They were from various dialect regions in Germany, with five speaking the Franconian dialect, three speaking other regional dialects, and six speaking only standard German. However, all of the talkers reported being proficient in standard German. American English talkers were recruited from the introductory psychology subject pool at the University of Texas at Austin, and were given credit towards partial fulfillment of course requirements. French and German talkers were paid $12/hour for their participation.

B. Speech materials

The three groups of talkers produced native-language vowels in a word or nonsense word context embedded in a language-appropriate sentence frame. Eleven American English vowels (/æ, ɛ, ɪ, eᶦ, i, ɑ, ʌ, ʊ, oᵘ, u, ɝ/) were produced in /b_t/ context within the sentence frame, "The large /*b*V̱*t*/ is green." Nine French vowels (/i, ɛ, a, ɑ, o, y, ɔ, œ, u/) were spoken in /p_t/ context within the sentence frame, "La grande /*p*V̱*t*/ est bleue," which translates as "The large /*p*V̱*t*/ is blue." Fifteen German vowels (7 short, /ɑ, ɛ, ɪ, ɔ, œ, ʊ, ʏ/ and eight long, /ɑː, ɛː, eː, iː, oː, øː, uː, ʏː/) were spoken in /b_tə/ context within the sentence frame "Matthias sagt /bVtən/," which translates as "Matthias says /bVtən/." Whereas the German vowel set corresponded to the entire German vowel inventory, excluding diphthongs, the American English and French sets were subsets of their respective inventories. The English set lacked the vowel /ɔ/ (which is usually not distinguished from /ɑ/ in Central Texas dialect) and the diphthongs /ɑᶦ/, /ɑᵘ/, and /ɔᶦ/; the French set lacked the close mid vowels /e/ and /ø/ (which are similar to their open mid counterparts) and all nasalized vowels.

C. Elicitation and recording procedure

Talkers were given an orthographic version of each sentence they were to produce (see the Appendix for orthographic representations of test words). They were then presented with a written question about the sentence that was designed to encourage them to place the focus on either the target word (i.e., the word containing the target vowel), or on the final adjective in the sentence. In English, for example, when talkers read the question, "The large WHAT is blue?," the target word represented new information for the hypothetical person who had asked the question (i.e., the target word was marked [+focus]). When talkers read the question, "The large /*b*V̱*t*/ is WHAT?," the target word represented given information for the hypothetical question asker (i.e., the target word was marked [−focus]). Talkers were instructed to produce the entire sentence when responding to the question, and their responses were recorded. They were also instructed to answer the questions at a normal rate of speech.[3] For a given vowel, talkers produced each focus condition before moving on to the next vowel. The order of focus condition was randomized between vowels, and the order of vowels was randomized between talkers. Productions by English and German speakers were digitally recorded at a sample rate of 22 050 Hz using a Shure SM48 microphone and a locally developed signal processing program (WAVAX) implemented on a PC. Utterances of French speakers were recorded using a Shure HW501 headworn dynamic microphone and a Sony Mini Disc MZ-R70 recorder and were then imported into WAVAX at the same sampling rate. All recordings were made either in a sound-attenuated booth or in a quiet room.

D. Acoustic measurements

Vowel tokens were analyzed acoustically using the Multi-Speech Signal Analysis Workstation, Model 3700, Version 2.4 (Copyright© 2001 Kay Elemetrics Corp). Temporal measurements were made from inspection of wave forms and wideband spectrograms (323 Hz bandwidth). Vowel durations were measured as the interval between the first and last significant glottal pulses of a given token. As noted earlier, one possible means of distinctiveness enhancement of vowels in [+focus] context is to increase systematic variation in vowel duration (e.g., between tense and lax vowels in English or between long and short vowels in German). Accordingly, for all three languages examined in this study, we calculated the variance in vowel durations for each talker and condition.

Frequencies of the first two formants (F1 and F2) were measured using pitch synchronous LPC (18 coefficients) with a 20 ms Blackman analysis window centered at each of three relative temporal locations—25%, 50%, and 75% of vowel duration. LPC derived formant tracks were superimposed on the spectrograms. In cases where the formant tracks reflected spurious peaks or where actual formants were missed, formant frequencies were marked manually using visual inspection of the spectrograms and of 512-point FFT spectra. The F1 and F2 measurements were used to determine the size of the vowel space and the amount of VISC for each talker.

E. Auditory modeling and measurements

Formant frequency measurements are very useful for making comparisons with other studies of vowels. However, the motivating idea of this study—sufficient contrast—must ultimately be evaluated in auditory, rather than purely acoustic, terms. Accordingly, an auditory model was used to produce auditory spectra of each vowel token at the 25%, 50%, and 75% relative time locations. These spectral data were analyzed using principal components analysis (PCA), and the first two extracted components (PC1 and PC2) served to define a two-dimensional space (analogous to an $F1 \times F2$ space) within which vowel space size and amount of VISC could be calculated.

The auditory spectra were plotted in units of loudness (Sones) at each 0.2 Bark step, for English and French, and at each 0.1 Bark step for German, between 0 and 16 Bark (corresponding to the frequency range between 0 and 3200 Hz). (The use of a smaller Bark step for German, corresponding to a smaller analysis bin, was necessitated by the relatively low fundamental frequency values of several male talkers, which would otherwise have produced more than one harmonic per analysis bin.) These auditory spectra were derived using a modified version of the auditory model described by Bladon and Lindblom (1981), which was itself adapted from Schroeder *et al.* (1979). This model was judged to be suitable for our purposes because it was based on human psychophysical measurements rather than on physical or electrophysiological data from animal listeners. Although the Bark frequency scale of the model was retained in the present application, we replaced the Bark-based auditory filters with filters having somewhat narrower equivalent rectangular bandwidth (ERB) values as described by Moore and Glasberg (1983) and Glasberg and Moore (1990).

The Bladon and Lindblom (1981) model operates on schematic vowel spectra each consisting of a list of harmonic frequencies and amplitudes. In order to apply the model to actual vowel signals, it is necessary to preprocess the signals, converting actual spectra into schematic spectra. This was done as follows: A Hanning window was applied to a selected temporal region of the vowel consisting of at least three glottal periods, and an FFT was calculated. (Hanning windowing results in a spectrum consisting of triangular shaped harmonics.) A set of frequency-domain template functions were created each consisting of a series of triangular pulses of unit height and spaced at equal frequency intervals, with the size of the frequency interval varying in very small steps across the template set. A given FFT spectrum was multiplied by each template function of the set, and the template yielding the highest product was used to estimate the harmonic frequencies of the FFT spectrum. The corresponding harmonic amplitudes were then determined directly from the FFT spectrum. Visual inspection of sample cases confirmed that this procedure yielded highly accurate results.

F. Calculating vowel space size and amount of VISC

For both the acoustic ($F1 \times F2$) and auditory ($PC1 \times PC2$) spaces, two methods were used to calculate the size of the vowel space. These methods are described here in terms that apply equally to the acoustic and auditory spaces. The first method was to calculate the area of the triangle (using Heron's formula) defined by the vowels /i/, /ɑ/, and /u/, with the analysis window centered at the 50% time point of each vowel trajectory. An advantage of this quantity is that it corresponds to a typical graphic representation of vowel space size in two dimensions. The second method was to calculate the average Euclidean distance between every pair of vowels in the entire vowel set. Relative to the first method, this one yields a more comprehensive measure of vowel dispersion. The quantities calculated by these methods are referred to here as size (area) and size (dispersion), respectively.

Two methods were also used to calculate the amount of VISC according to both the acoustic and auditory measures. The first was to sum the Euclidean distances traversed by a given vowel (in the $F1 \times F2$ space or the $PC1 \times PC2$ space) between the 25% and 50% time points and between the 50% and 75% time points and then to average these sums across the entire vowel set. This quantity, which is referred to as VISC (magnitude), approximates the degree of spectral change across the vowel nucleus while limiting somewhat the contribution of the initial and final consonants to spectral dynamics. It can reasonably be viewed as a valid measure of the amount of VISC. However, by itself this quantity does not provide for an adequate test of the notion of sufficient contrast. The reason is that there is no guarantee that a greater amount of VISC, defined in this way, will correspond to greater acoustic or auditory distinctiveness among vowels.

The second method for calculating the amount of VISC was an attempt to remedy this limitation. It was analogous to the size (dispersion) method for calculating vowel space size. For each stimulus, the signed differences in F1 and F2 (or PC1 and PC2) were calculated between the 25% and 50% time points and between the 50% and 75% time points. For each of the two time intervals, every vowel token was represented in a space whose dimensions corresponded to these signed differences (i.e., $F1_{diff} \times F2_{diff}$, or $PC1_{diff} \times PC2_{diff}$), and the average Euclidean distance between every pair of vowel tokens was determined. Finally, these two averages were themselves averaged across the two time intervals. This quantity, which is referred to as VISC (dispersion), yields a measure of vowel distinctiveness that is independent of size (dispersion).

III. RESULTS

For each language and for each measure, a 2 (focus condition) ×2 (sex of talker) repeated-measures ANOVA was performed. In the case of German, ANOVAs for size (area) were performed on the short vowel and long vowel sets separately, whereas those for size (dispersion) and all other measures were performed on the entire vowel set.

A. Vowel space size

1. Acoustic space: *F*1 × *F*2

a. Size (area): For all three languages, there was a significant effect of focus condition (English: $[F(1,10)=10.70, p<0.01, \eta_p^2=0.517]$, French: $[F(1,10)=15.86, p<0.005, \eta_p^2=0.613]$, German long: $[F(1,12)=13.46, p<0.005, \eta_p^2=0.529]$ and German short: $[F(1,12)=15.79, p<0.005, \eta_p^2=0.568]$), with [+focus] vowels having a larger size (area) than [−focus] vowels. Sex of talker was not significant in either English or French but was significant for both German long $[F(1,12)=48.58, p<0.001, \eta_p^2=0.802]$ and short $[F(1,12)=21.04, p<0.005, \eta_p^2=0.637]$ vowels. Figures 1–4 show the effects of focus condition and sex of talker on the size of the acoustically defined vowel space for each of the three languages.

b. Size (dispersion): Figure 5(a) shows the effect of focus condition on the acoustically defined measure of size (dispersion). This effect was significant for English $[F(1,10)=33.10, p<0.001, \eta_p^2=0.768]$, French $[F(1,10)=41.61, p<0.001, \eta_p^2=0.806]$, and German $[F(1,12)=37.14, p<0.001, \eta_p^2=0.756]$, with [+focus] vowels having greater size (dispersion) than [−focus] vowels. The effect of sex of talker was significant only in German ($[F(1,12)=19.12, p<0.005, \eta_p^2=0.614]$), with German females having greater size (dispersion) than German males.

2. Auditory space: PC1 × PC2

a. Size (area): For English and French, there was a significant effect of focus condition on the size of the auditorily defined vowel space ($[F(1,10)=10.37, p<0.01, \eta_p^2=0.509]$ and $[F(1,10)=9.78, p<0.05, \eta_p^2=0.499]$, respectively), with the [+focus] vowels having a larger vowel space than [−focus] vowels. German short $[F(1,12)=19.91, p<0.005, \eta_p^2=0.624]$, but not long vowels ($p=0.072$), were produced with significantly more auditory separation in the [+focus] condition. Sex of talker was not a significant variable in any of the three languages. The results patterned very similarly for the acoustically and auditorily defined vowel spaces. Table I displays correlations, for each of the three languages, between the first and second principal components of the auditorily defined space and F1 and F2.

b. Size (dispersion): Figure 5(b) displays the effect of focus condition on the auditorily defined measure of size (dispersion). For all three languages, [+focus] vowels showed greater size (dispersion) than [−focus] vowels (English: $[F(1,10)=49.76, p<0.001, \eta_p^2=0.833]$, French $[F(1,10)=15.04, p<0.005, \eta_p^2=0.601]$, German $[F(1,12)=15.14, p<0.005, \eta_p^2=0.558]$. As in the case of the auditorily defined size (area) measure, sex of talker was not a significant variable for any of the three languages.

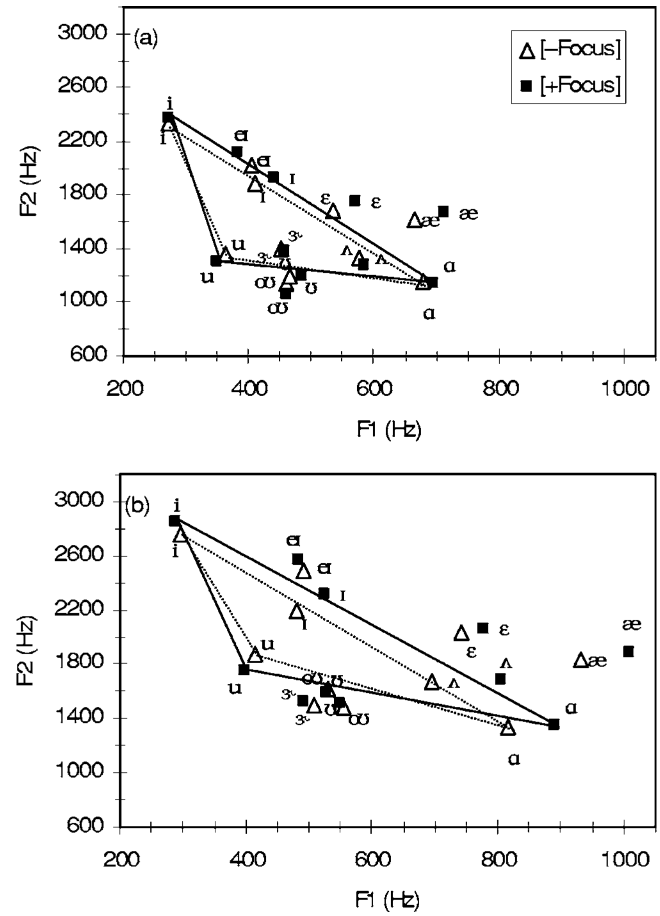

FIG. 1. Effect of focus condition on size (area) for English vowels: (a) males, (b) females. Dotted and solid lines represent the /i/-/ɑ/-/u/ vowel triangles for the [−focus] and [+focus] conditions, respectively.

B. VISC

1. Acoustic space: F1 × F2 for VISC (magnitude); $F1_{diff} \times F2_{diff}$ for VISC (dispersion)

a. VISC (magnitude): Figure 6(a) shows the effect of focus condition on VISC (magnitude) for the three languages. Focus condition was a significant variable only for

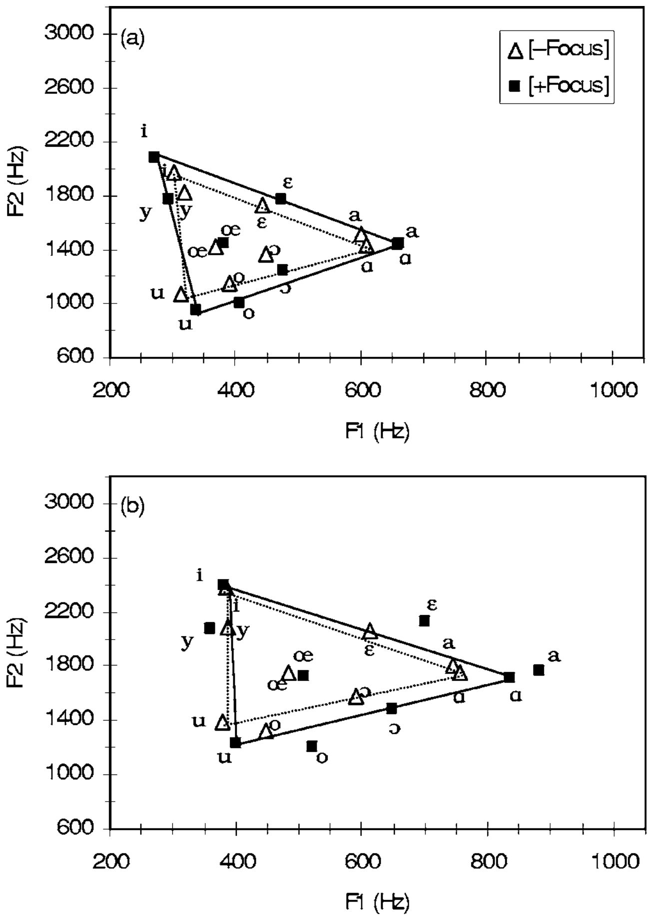

FIG. 2. Effect of focus condition on size (area) for French vowels: (a) males, (b) females. Dotted and solid lines represent the /i/-/ɑ/-/u/ vowel triangles for the [−focus] and [+focus] conditions, respectively.

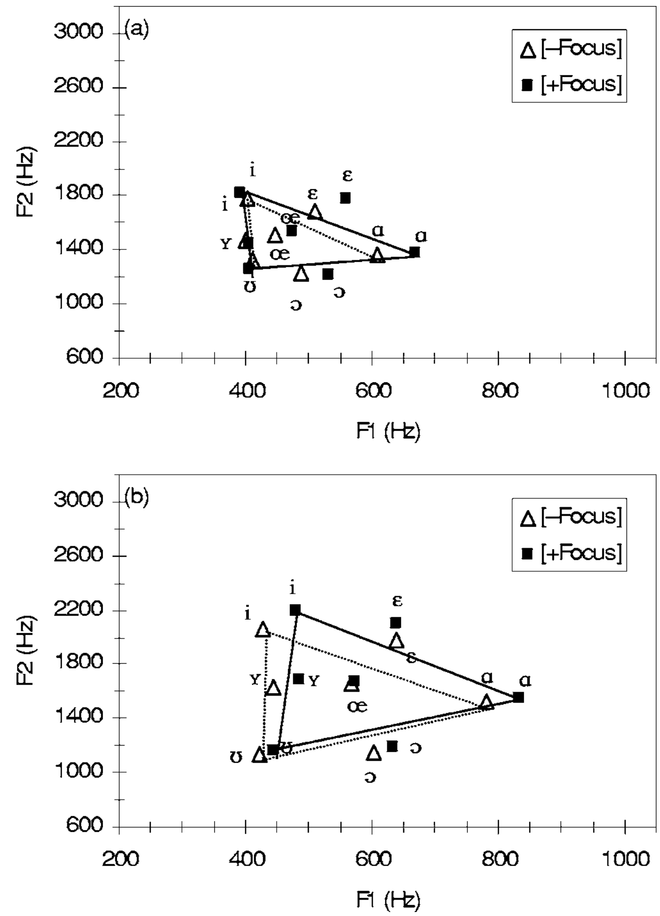

FIG. 3. Effect of focus condition on size (area) for German short vowels: (a) males, (b) females. Dotted and solid lines represent the /i/-/ɑ/-/u/ vowel triangles for the [−focus] and [+focus] conditions, respectively.

German [$F(1,12)=14.54$, $p<0.005$, $\eta_p^2=0.548$], with [+focus] vowels showing greater VISC (magnitude) than [−focus] vowels. For all three languages, female talkers showed significantly greater VISC (magnitude) than male talkers (English: [$F(1,10)=13.57$, $p<0.005$, $\eta_p^2=0.576$], French: [$F(1,10)=14.87$, $p<0.005$, $\eta_p^2=0.598$], and German: [$F(1,12)=24.76$, $p<0.001$, $\eta_p^2=0.674$]). There were no significant interactions between focus condition and sex of talker for any of the three languages.

b. VISC (dispersion): Figure 6(b) exhibits the effect of focus condition on VISC (dispersion) for the three languages. As in the case of VISC (magnitude), German was the only language to show significantly greater VISC (dispersion) for [+focus] vowels than [−focus] vowels [$F(1,12)=11.87$, $p<0.01$, $\eta_p^2=0.497$]. For each of the three languages, there was a significant effect of sex of talker on the measure of VISC (dispersion) (English: [$F(1,10)=26.12$, $p<0.001$, $\eta_p^2=0.723$], French: [$F(1,10)=10.36$, $p<0.01$, $\eta_p^2=0.509$], and German: [$F(1,12)=33.64$, $p<0.001$, $\eta_p^2=0.737$]), with females showing greater acoustic separation than males. There were no significant interactions between focus condition and sex of talker for any of the languages.

2. Auditory space: PC1×PC2 for VISC (magnitude); $PC1_{diff}\times PC2_{diff}$ for VISC (dispersion)

a. VISC (magnitude): The effect of focus condition on the auditorily defined measure of VISC (magnitude) is shown in Fig. 6(c). English, French, and German all showed greater VISC for [+focus] vowels than [−focus] vowels ([$F(1,10)=111.11$, $p<0.001$, $\eta_p^2=0.917$], [$F(1,10)=16.41$, $p<0.005$, $\eta_p^2=0.621$], and [$F(1,12)=15.09$, $p<0.005$, $\eta_p^2=0.557$], respectively). None of the languages showed an effect of sex of talker.

As was true for size (area), correlations between the first two principal components of the auditorily defined space and F1 and F2 were generally significant. These are shown in Table I.

b. VISC (dispersion): Figure 6(d) displays the effect of focus condition on the auditorily defined measure of VISC (dispersion). All three languages showed a significant effect of focus condition (English [$F(1,10)=77.05$, $p<0.001$, $\eta_p^2=0.885$], French [$F(1,10)=9.18$, $p<0.05$, $\eta_p^2=0.478$], and German [$F(1,12)=14.10$, $p<0.005$, $\eta_p^2=0.540$]), with [+focus] vowels showing greater separation than [−focus] vowels. There was no significant effect of sex of talker for any of the languages.

Note that for both VISC (magnitude) and VISC (dispersion), the auditorily defined measures yielded more cross-linguistically consistent effects of focus condition than the acoustically defined measures.

In contrast, the auditorily defined measures of both vowel space size and VISC showed, if anything, less sensitivity to the sex of talker variable. This can perhaps be attributed to the use of the Bark scale in the auditory represen-

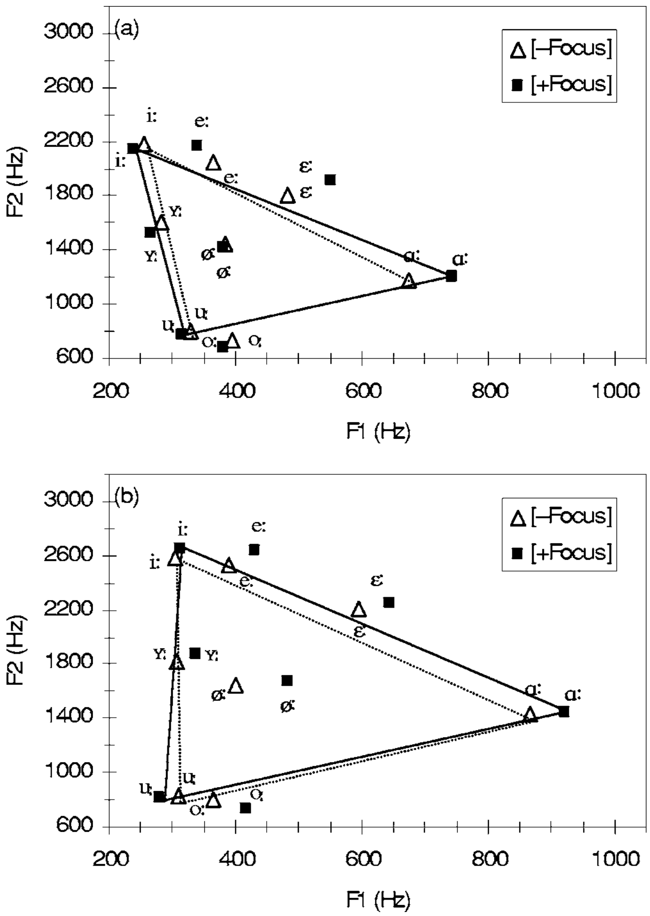

FIG. 4. Effect of focus condition on size (area) for German long vowels: (a) males, (b) females. Dotted and solid lines represent the /i:/-/ɑ:/-/u:/ vowel triangles for the [−focus] and [+focus] conditions, respectively.

tations of vowels, the effect of which is partly to normalize frequency differences between male and female talkers (Syrdal and Gopal, 1986).

Although the measures of VISC (magnitude) and VISC (dispersion) are logically independent, they should be highly correlated provided that vowel tokens are widely distributed in the $F1_{diff} \times F2_{diff}$ space or the $PC1_{diff} \times PC2_{diff}$ space. Table II shows these correlations for each language and for both the acoustic and auditory measures.

C. Duration variance

Neither English nor French showed significant effects of focus condition on vowel duration variance. However, in German there was more duration variance for [+focus] vowels than for [−focus] vowels [$F(1,12)=17.01$, $p<0.005$, $\eta_p^2=0.586$]. There was no significant effect of sex of talker on duration variance in any of the languages. Figure 7 shows the effects of focus condition on duration variance for each of the three languages.

TABLE I. Correlations between auditory principal components and formant frequencies. The asterisks represent that correlation is significant at the 0.01 level (2-tailed).

		Vowel space		VISC	
		PC1	PC2	PC1	PC2
English	F1	0.046	0.471*	0.446*	0.018
English	F2	0.184*	−0.450*	−0.472*	0.171*
French	F1	0.662*	0.289*	0.651*	0.275*
French	F2	−0.359*	0.422*	−0.386*	0.401*
German	F1	0.627*	0.159*	0.623*	0.129*
German	F2	−0.261*	0.452*	−0.295*	0.438*

D. Comparing the English tense/lax distinction and the German long/short distinction

In the analyses described above, each language was treated separately owing to differences among English, French, and German in the size and character of the vowel inventories. However, for a subset of the measurements, it is instructive to perform additional analyses in which language is a factor. This subset includes data for all seven long/short pairs of vowels in German (/ɑ:, ɑ/, /e:, ɛ/, /i:, ɪ/, /o:, ɔ/, /ø:, œ;/, /u:, ʊ/, /ʏ:, ʏ/) as well as for four tense/lax vowel pairs in America English (/ɑ, ʌ/, /e[ɪ], ɛ/, /i, ɪ/, /u, ʊ/). The members of each of these vowel pairs are potentially confusable because they occupy similar or adjacent regions in $F1 \times F2$ space, although in both languages (and especially in German) this problem is mitigated because of durational differences.

Three sets of measurements were analyzed using 2 (focus conditions) × 2 (language) repeated-measures ANOVAs. (For the purpose of these analyses, the data were collapsed over the sex of talker variable.) The first two sets of measurements were equivalent to the auditory versions of size (dispersion) and VISC (dispersion), respectively, except that the stimulus tokens being compared did not comprise the entire vowel set but instead were restricted to the two mem-

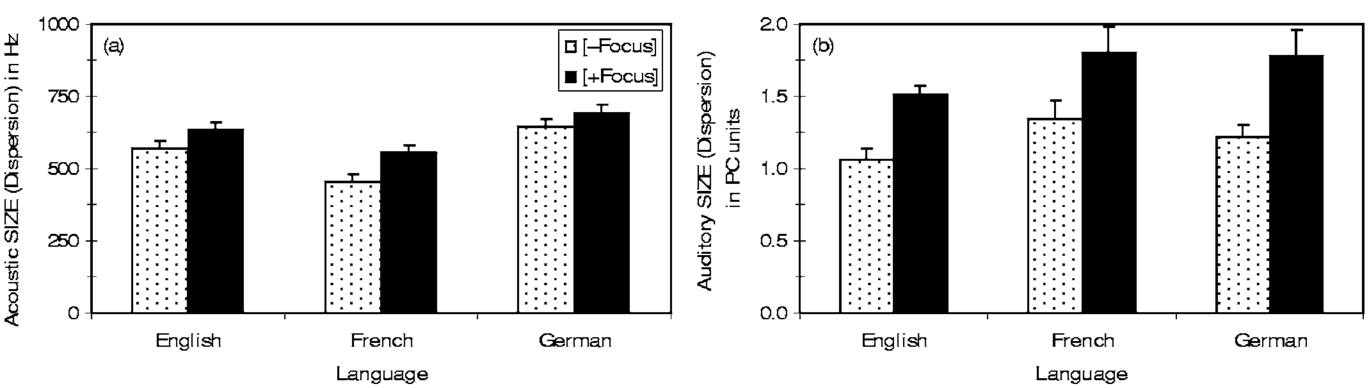

FIG. 5. Effect of focus condition on the (a) acoustically defined and (b) auditorily defined measures of size (dispersion) for English, French, and German. PC units refer to unit distances in the $PC1 \times PC2$ space. Error bars represent the standard error of the mean.

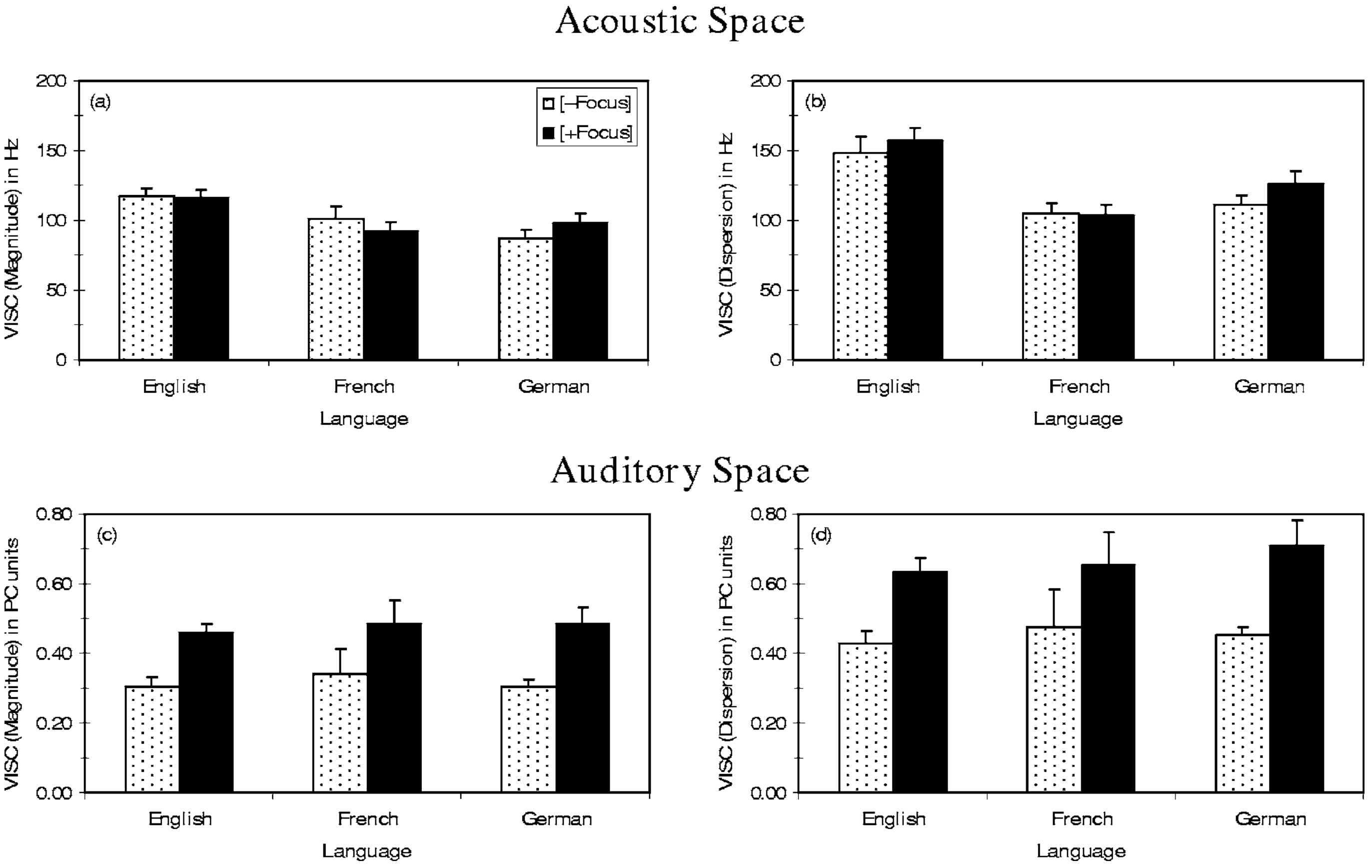

FIG. 6. Effect of focus condition on the acoustically defined measures of (a) VISC (magnitude) and (b) VISC (dispersion) and on the auditorily defined measures of (c) VISC (magnitude) and (d) VISC (dispersion), for English, French, and German. For the size (dispersion) measure, PC units refer to unit distances in the PC1 × PC2 space; for the VISC (dispersion) measure, PC units refer to unit distances in the $PC1_{diff} \times PC2_{diff}$ space. Error bars represent the standard error of the mean.

bers of each long/short or tense/lax pair. The third set of measurements consisted of the duration ratios between long/short or tense/lax pair members.

Table III shows the mean values of size (dispersion), VISC (dispersion), and duration ratio between English tense/lax and German long/short vowel pair members in [+focus] and [−focus] conditions. There were significant effects of focus condition for the modified size (dispersion) [$F(1,24)=31.89$, $p<0.001$, $\eta_p^2=0.571$] and VISC (dispersion) [$F(1,24)=27.13$, $p<0.001$, $\eta_p^2=0.531$] measures, with both yielding larger values in [+focus] condition. For neither measure was there a significant effect of language or a significant focus condition × language interaction.

For the duration ratio measure, there were significant main effects of focus condition [$F(1,24)=4.81$, $p<0.05$, $\eta_p^2=0.167$] and language [$F(1,24)=102.43$, $p<0.001$, $\eta_p^2=0.810$], with vowel pairs in [+focus] condition and in German showing greater duration ratios, but there was no interaction between these variables. Analyses of simple effects showed that focus condition was a significant variable in German [$F(1,13)=8.04$, $p<0.05$, $\eta_p^2=0.382$], but not in English ($p=0.497$). The larger duration ratios for [+focus] vowels in German help to account for the increased overall duration variance reported earlier for German vowels in [+focus] condition. Although the tense/lax duration ratios in English were much smaller than the long/short duration ratios in German, the former were nevertheless significantly greater than 1.0 ($p<0.001$) in both [+focus] and [−focus] conditions.

It is noteworthy that the relative effect of focus condition was considerably greater for the two dispersion measures than for the duration ratio measure. This was no less true in German than in English.

TABLE II. Correlations between VISC (magnitude) and VISC (dispersion) in $F1_{diff} \times F2_{diff}$ and $PC1_{diff} \times PC2_{diff}$ spaces for English, French, and German. The asterisks represent that correlation is significant at the 0.01 level (2-tailed).

	Acoustic ($F1_{diff} \times F2_{diff}$ Space)	Auditory ($PC1_{diff} \times PC2_{diff}$ Space)
English	0.908*	0.964*
French	0.859*	0.990*
German	0.966*	0.994*

IV. DISCUSSION

The present study examined several possible acoustic and auditory correlates of vowels in [+focus] context produced by talkers of American English, French, and German. Consistent with Lindblom's theory of adaptive dispersion, each language group enhanced the distinctiveness of vowels in the portion of an utterance used to signal new (as opposed to given) information, and, as allowed by the principle of sufficient contrast, the three language groups used varying

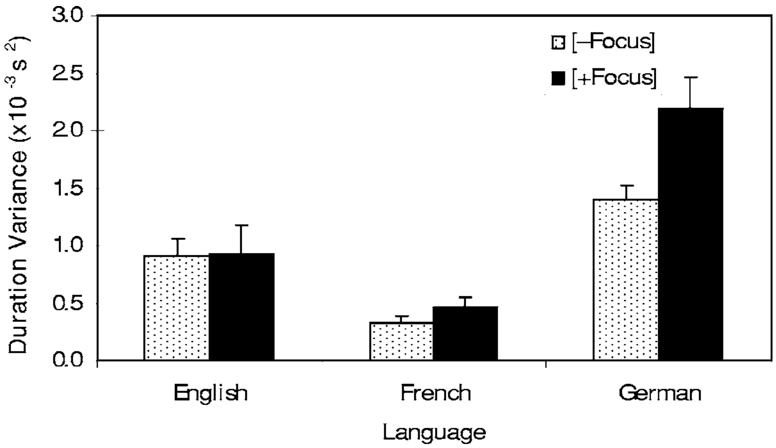

FIG. 7. Effect of focus condition on duration variance in English, French, and German. Error bars represent the standard error of the mean.

means of enhancing distinctiveness. In producing vowels in [+focus] context, all three language groups increased spectral differences among vowels, but only the German talkers increased vowel duration differences. Table IV lists the measures for which the effect of focus condition was significant in each language.

The acoustic and auditory versions of the spectral measures generally yielded similar patterns of results; however, the auditory versions yielded a higher number of significant effects of focus condition on the measures of VISC. We interpret these results as a validation of our auditory measures of spectral distance. Although the vowel space is most commonly defined acoustically, namely, in terms of the frequencies of the first several formants, there are several advantages in using an auditorily defined space. First, as noted earlier, the notion of sufficient contrast is more directly evaluated in terms of auditory rather than acoustic distance. Second, automatic formant tracking procedures are error-prone and often need to be supplemented with manual editing. In contrast, the auditory space used in the present study is derived entirely by algorithmic means. Third, there is little reason to believe that, unlike artificial systems, the human auditory system some how manages to circumvent the inherent difficulties in extracting formant frequencies reliably. It seems more plausible to assume that the auditory system extracts parameters that are rich in phonetic information but that can be computed algorithmically from the speech signal. Among possible candidates for such parameters, we would include stimulus values within a PCA-defined auditory space such as that used in the present study.[4] Previous factor analyses of auditory spectra have yielded principal components that account well for the variance among vowel sounds (Plomp *et al.*, 1967; Klein *et al.*, 1970) and that are reasonably highly correlated with formant frequencies (Nearey and Kiefte, 2003). Such correlations were also found in this study. Factor analytic approaches in these cases may thus be described as yielding implicit information about formant values without requiring actual formant tracking.

Finally, we return to the topic of cross-language variation in the means by which vowel distinctiveness is enhanced in [+focus] contexts. One possible account of this variation is based on the distinction between phonologically contrastive and noncontrastive, or secondary, properties. Consider, for example, the case of vowel duration differences. In German, such differences are used to signal phonological contrasts between long and short vowels of similar quality (e.g., /ɑː/ versus /ɑ/). However, in English, vowel duration differences are generally viewed as a type of noncontrastive variation that is correlated with and conditioned by some primary phonological contrast (e.g., tense versus lax vowels and nonlow versus low vowels). As in other languages that make contrastive use of vowel length (Heldner and Strangert, 2001; de Jong and Zawaydeh, 2002; Smiljanic, 2004), duration ratios of German long/short vowel pairs were significantly increased in [+focus] context. However, no such change occurred for English tense/lax vowel pairs. This pat-

TABLE III. Mean auditory size (dispersion), VISC (dispersion), and duration ratio for English tense/lax and German long/short vowel pairs. For the size (dispersion) measure, PC units refer to unit distances in the PC1 ×PC2 space; for the VISC (dispersion) measure, PC units refer to unit distances in the $PC1_{diff} \times PC2_{diff}$ space.

	English		German	
	[+Focus]	[−Focus]	[+Focus]	[−Focus]
Size (dispersion) in PC units	1.056	0.761	1.266	0.719
VISC (dispersion) in PC units	0.578	0.396	0.688	0.396
Duration ratio	1.238	1.183	2.126	1.981

TABLE IV. Significant effects of focus condition on distinctiveness enhancement in vowels. * represents significant at the 0.05 level; ** significant at the 0.01 level; *** significant at the 0.001 level.

	English	French	German
Acoustic size(area)	**	**	**
Acoustic size(dispersion)	***	***	***
Auditory size(area)	**	**	Long=ns, Short***
Auditory size(dispersion)	***	**	**
Acoustic VISC(magnitude)	ns	ns	**
Acoustic VISC(dispersion)	ns	ns	**
Auditory VISC(magnitude)	***	**	**
Auditory VISC(dispersion)	***	*	**
Duration variance	ns	ns	***

tern of results might suggest that in [+focus] context talkers exaggerate only those vowel differences that are phonologically contrastive.

However, this account of cross-language differences in vowel contrast enhancement faces several difficulties. First, some noncontrastive properties of phonemes do appear to be subject to enhancement in [+focus] contexts. In modern theories of distinctive features (for a review, see Diehl and Lindblom, 2004), VISC has not generally been recognized as being phonologically contrastive. It is more likely to be viewed as a secondary correlate of certain primary vowel features (e.g., tenseness or height). Yet, in the present study, talkers from all three language groups exhibited increased VISC (magnitude) or increased VISC (dispersion) in [+focus] context, by either the acoustic measure, the auditory measure, or both. Analogously, it has been reported that duration differences of English stressed vowels before [+voice] and [−voice] stop consonants are enhanced in [+focus] context (de Jong, 2004), even though such differences have typically been viewed as a secondary correlate of the syllable-final [voice] distinction. Thus, status as a phonologically contrastive feature of a language does not appear to be a necessary condition for enhanced distinctiveness in utterance focus.

Another difficulty in appealing to the contrastive/noncontrastive distinction to account for cross-language differences in vowel contrast enhancement is that the distinction itself is difficult to apply in practice and may be theoretically problematic. For example, German vowels differing in length have been variously described by linguists as tense/lax contrasts or as long/short contrasts (Hawkins, 1992). Phonetically, the members of these vowel pairs differ both in spectral properties and in duration (Fischer-Jørgensen, 1990), and it is unclear which of these differences should be viewed as contrastive and which as noncontrastive (or secondary). Because phonological features may be quite abstract, mapping onto diverse phonetic correlates each having perceptual relevance (Kingston and Diehl, 1994), the assignment of contrastive status to one correlate and secondary status to others may be quite arbitrary.

An alternative hypothesis is that the amount of contrast enhancement of a vowel property in [+focus] context is positively related to the between-category variation of that property in [−focus] context. Although the spectral measures did not provide evidence for or against this hypothesis (because the language groups did not differ very much according to these measures in either the [−focus] or the [+focus] conditions), the duration variance and duration ratio measures did yield results consistent with the hypothesis. Duration variance among [−focus] vowels was much larger in German than in English and French, and significant enhancement of duration variance in [+focus] context occurred only for German.[5] Similarly, duration ratios between German long/short vowel pairs were much larger than those between English tense/lax vowel pairs in [−focus] condition, and only German vowels exhibited significant enhancement of duration ratios in [+focus] condition. Some previous cross-language results also appear to be consistent with the above hypothesis. For example, de Jong and Zawaydeh (2002) and de Jong (2004) found that vowel duration differences conditioned by a following [voice] distinction are greater in English than in Arabic (see also Port *et al.*, 1980, and Flege and Port, 1981), and that [+focus] contexts yield a reliable enhancement of these differences only in English. A tendency to enhance properties of phonemes that show greater between-category variance may be viewed as yet another instance of the positive relationship, predicted by the theory of adaptive dispersion, between utterance clarity and information content.

ACKNOWLEDGMENTS

This work was supported by research grants Nos. R01 DC00427-13, -14, -15 from the National Institute of Health, National Institute on Deafness and Other Communication Disorders, to the fifth author (R.L.D.) Special thanks to Juan Segui and Ludovic Ferrand of the CNRS and Université René Descartes, Laboratoire de Psychologie Expérimental, Paris, France, for providing facilities to conduct the French portion of the study. Some of these findings were reported at the 143rd Meeting of the Acoustical Society of America, Pittsburgh, June 4, 2002 and at the 146th Meeting of the Acoustical Society of America, Austin, November 11, 2003. The authors thank Ann Bradlow for very helpful comments on earlier versions of this paper.

APPENDIX

The following orthographic representations of test words were used:

English material

bat bæt	*bet* bɛt	*bit* bɪt	*bait* be^{ɪ}t	*beet* bit
baht bat **	*butt* bʌt	*bout* bʊt rhymes w/‘put’ **	*boat* bo^{ʊ}t	*boot* but
bert bɝt **				

French material

pites pit **	*pète* pɛt ‘to break’ (present tense)	*patte* pat ‘paw’	*pâte* pɑt ‘pasta’	*peaute* pot **
pute pyt ‘hooker’	*pote* pɔt ‘friend’ (colloquial)	*peute* pœt **	*poute* put **	

German material

Batten bɑtən **	*Betten* bɛtən ‘beds’	*Bitten* bɪtən ‘to ask’	*Botten* bɔtən **	*Bötten* bœtən **
Butten bʊtən **	*Bütten* bʏtən ‘tubs’	*Baten* bɑ:tən ‘asked’	*Bäten* bɛ:tən ‘asked’ (subj)	*Beten* be:tən ‘pray’
Bieten bi:tən ‘to offer’	*Boten* bo:tən ‘boats’	*Böten* bɵ:tən **	*Buten* bu:tən **	*Büten* bʏ:tən **

**Nonsense words.

[1]Under noisy conditions, clear speech has also been found to be more intelligible than reduced speech, other things being equal, for normal hearing adults (Tolhurst, 1957) and children with learning disabilities (Bradlow *et al.*, 2003).

[2]Apart from signaling new information, [+focus] may also be used to highlight thematic relations between a constituent and its context (Nooteboom and Kruyt, 1987).

[3]For the American English and French groups but not the German group, the same questions were also answered using fast and slow rates of utterance. For a given vowel, talkers produced each focus condition at every rate before moving on to the next vowel. The order of focus condition and utterance rate was randomized between vowels. Because focus condition was the key variable of interest in the present study and because the German group produced utterances only at the normal rate, we include here only results for the normal utterance rate. In general, the pattern of results obtained for the normal utterance rate was very similar to the patterns for the other two utterance rates.

[4]The demand for efficient neural coding of signals suggests that some form of factor analysis (e.g., PCA or independent components analysis) may be implicit in the design of sensory/perceptual systems (Field, 1987; Simoncelli and Olshausen, 2001; Lewicki, 2002).

[5]A potential concern with the participant selection in the present study is that, unlike the American English and French groups, the German group was currently visiting or residing in Autsin, Texas, rather than living in their native language community. However, if exposure to English affected the German vowel productions, this was clearly not sufficient to eliminate differences between English and German modes of contrast enhancement.

Altman, G., and Carter, D. M. (**1989**). "Lexical stress and lexical discriminability: Stressed syllables are more informative, but why?," Comput. Speech Lang. **3**, 265–275.

Anderson, A. H., Bard, E. G., Sotillo, C., Newlands, A., and Doherty-Sneddon, G. (**1997**). "Limited visual control of the intelligibility of speech in face-to-face dialogue," Percept. Psychophys. **59**, 580–592.

Baart, J. J. G. (**1987**). "Focus, syntax, and accent placement," Ph.D. thesis, Leyden University.

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., and Newlands, A. (**2000**). "Controlling the intelligibility of referring expressions in dialogue," J. Mem. Lang. **42**, 1–22.

Billerey-Mosier, R. (**2000**). "Segment realization in relation to a carrier word's uniqueness point," J. Acoust. Soc. Am. **108**, 2479–2480.

Bladon, R. A. W., and Lindblom, B. (**1981**). "Modeling the judgment of vowel quality differences," J. Acoust. Soc. Am. **69**, 1414–1422.

Bolinger, D. (**1989**). *Intonation and Its Uses* (Stanford University Press, Stanford).

Bradlow, A. R., Kraus, N., and Hayes, E. (**2003**). "Speaking clearly for children with learning disabilities: Sentence perception in noise," J. Speech Lang. Hear. Res. **46**, 80–97.

Browman, C. P., and Goldstein, L. (**1995**). "Gestural syllable position effects in American English," in *Producing Speech: Contemporary Issues. For Katherine Harris*, edited by F. Bell-Berti and L. J. Raphael (AIP Press, New York), pp. 19–33.

Byrd, D. (**1994**). "Articulatory timing in English consonant sequences," Ph.D. thesis, University of California, Los Angeles.

Campbell, N. (**1995**). "Loudness, spectral tilt, and perceived prominence in dialogues," 13th International Congress of Phonetic Sciences (Stockholm, Sweden), pp. 676–679.

Cho, T., and Keating, P. (**2001**). "Articulatory strengthening at the onset of prosodic domains in Korean," J. Phonetics **28**, 155–190.

Cho, T., and McQueen, J. M. (**2005**). "Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress," J. Phonetics **33**, 121–157.

Chomsky, N., and Halle, M. (**1968**). *The Sound Pattern of English* (Harper and Row, 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.

de Jong, K. (**2004**). "Stress, lexical focus, and segmental focus in English: patterns of variation in vowel duration," J. Phonetics **32**, 493–516.

de Jong, K., and Zawaydeh, B. (**2002**). "Comparing stress, lexical focus, and segmental focus: patterns of variation in Arabic vowel duration," J. Phonetics **30**, 53–75.

Diehl, R. L., and Lindblom, B. (**2004**). "Explaining the structure of feature and phoneme inventories: The role of auditory distinctiveness," in *Speech Processing in the Auditory System*, edited by S. Greenberg, W. A. Ainsworth, A. N. Popper, and R. R. Fay (Springer, New York), pp. 101–162.

Eefting, W. (**1991**). "The effect of 'information values' and 'accentuation' on the duration of Dutch words, syllables, and segments," J. Acoust. Soc. Am. **89**, 412–424.

Field, D. J. (**1987**). "Relations between the statistics of natural images and the response properties of cortical cells," J. Opt. Soc. Am. A **4**, 2379–2394.

Fischer-Jørgensen, E. (**1990**). "Intrinsic F_0 in tense and lax vowels with special reference to German," Phonetica **47**, 99–140.

Flege, J. E., and Port, R. F. (**1981**). "Cross-language phonetic interference: Arabic to English," Lang Speech **24**, 125–146.

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

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory shapes from notched-noise data," Hear. Res. **47**, 103–138.

Gussenhoven, C. (**1983**). "Focus, mode, and the nucleus," J. Ling. **19**, 377–417.

Hawkins, J. A. (**1992**). "German," in *International Encyclopedia of Linguistics*, edited by W. Bright (Oxford University Press, New York), pp. 61–69.

Heldner, M. (**2003**). "On the reliability of overall intensity and spectral emphasis as acoustic correlates of focal accents in Swedish," J. Phonetics **31**, 39–62.

Heldner, M., and Strangert, E. (**2001**). "Temporal effects of focus in Swedish," J. Phonetics **29**, 329–361.

Heldner, M., Strangert, E., and Deschamps, T. (**1999**). "A focus detector using overall intensity and high frequency emphasis," 14th International Congress of Phonetic Sciences (San Francisco, CA).

Hillenbrand, J. M., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099–3111.

Hillenbrand, J. M., and Nearey, T. M. (**1999**). "Identification of resynthesized /hVd/ utterances: Effects of formant contour," J. Acoust. Soc. Am. **105**, 3609–3523.

Hunnicutt, S. (**1985**). "Intelligibility vs redundancy—conditions on dependency," Lang Speech **28**, 47–56.

Jakobson, R., Fant, C. G. M., and Halle, M. (**1963**). *Preliminaries to Speech Analysis* (MIT Press, Cambridge).

Jun, S.-A. (**1993**). "The phonetics and phonology of Korean prosody," Ph.D. thesis, Ohio State University.

Keating, P., Cho, T., Fougeron, C., and Hsu, C. (**1999**). "Domain-initial articulatory strengthening in four languages," UCLA Working Papers in Linguistics **97**, 139–156.

Kingston, J., and Diehl, R. L. (**1994**). "Phonetic knowledge," Language **70**, 419–454.

Klein, W., Plomp, R., and Pols, L. C. W. (**1970**). "Vowel spectra, vowel spaces and vowel identification," J. Acoust. Soc. Am. **48**, 999–1009.

Knoll, M., and Uther, M. (**2004**). "Motherese and Chinese: Evidence of acoustic changes in speech directed at infants and foreigners," J. Acoust. Soc. Am. **116**, 2522.

Ladd, D. R. (**1980**). "The structure of intonational meaning," Ph.D. thesis, Indiana University.

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

Ladd, D. R., and Cutler, A. (**1983**). "Introduction: models and measurements in the study of prosody," in *Prosody: Models and Measurements*, edited by A. Cutler and D. R. Ladd (Springer-Verlag, Berlin), pp. 39–52.

Lane, H., and Tranel, B. (**1971**). "The Lombard sign and the role of hearing in speech," J. Speech Hear. Res. **14**, 677–709.

Lane, H., Tranel, B., and Sisson, C. (**1970**). "Regulation of voice communication by sensory dynamics," J. Acoust. Soc. Am. **47**, 618–624.

Lehiste, I. (**1970**). *Suprasegmentals* (MIT Press, Cambridge).

Lewicki, M. S. (**2002**). "Efficient coding of natural sounds," Nat. Neurosci. **5**, 356–363.

Lieberman, P. (**1963**). "Some effects of semantic and grammatical context on the production and perception of speech," Lang Speech **6**, 172–187.

Lindblom, B. (**1986**). "Phonetic universals in vowel systems," in *Experimental Phonology*, edited by J. J. Ohala and J. J. Jaeger (Academic, Orlando, FL), pp. 13–44.

Lindblom, B. (**1990**). "Explaining phonetic variation: A sketch of the H&H theory," in *Speech Production and Speech Modeling*, edited by W. J. Hardcastle and A. Marchal (Kluwer, Dordrecht), pp. 403–439.

Lombard, E. (**1911**). "Le signe d'elevation de la voix," Annales Maladies de l'Oreille, Larynx, Nez et Pharynx **37**, 101–119.

Luce, P. (**1986**). "Neighborhoods of words in the mental lexicon," Ph.D. thesis, Indiana University.

Lyberg, B. (**1979**). "Final lengthening—partly a consequence of restrictions on the speed of fundamental frequency change?," J. Phonetics **7**, 187–196.

Moore, B. C. J., and Glasberg, B. R. (**1983**). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," J. Acoust. Soc. Am. **74**, 750–753.

Nearey, T. M., and Assmann, P. F. (**1986**). "Modeling the role of inherent spectral change in vowel identification," J. Acoust. Soc. Am. **80**, 1297–1308.

Nearey, T. M., and Kiefte, M. (**2003**). "Comparison of several proposed perceptual representations of vowel spectra," 15th International Congress of Phonetic Sciences (Barcelona, Spain), pp. 1005–1008.

Nooteboom, S. G. (**1972**). "Production and perception of vowel duration," Ph.D. thesis, University of Utrecht.

Nooteboom, S. G., and Kruyt, J. G. (**1987**). "Accents, focus distribution, and the perceived distribution of given and new information: An experiment," J. Acoust. Soc. Am. **82**, 1512–1524.

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

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1985**). "Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech," J. Speech Hear. Res. **28**, 96–103.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1986**). "Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech," J. Speech Hear. Res. **29**, 434–446.

Plomp, R., Pols, L. C. W., and Van der Geer, J. P. (**1967**). "Dimensional analysis of vowel spectra," J. Acoust. Soc. Am. **41**, 707–712.

Pollack, I., and Pickett, J. (**1963**). "The intelligibility of excerpts from conversation," Lang Speech **6**, 165–171.

Port, R. F., Al-Ani, S., and Maeda, S. (**1980**). "Temporal compensation and universal phonetics," Phonetica **37**, 235–252.

Redford, M. A., and Diehl, R. L. (**1999**). "The relative perceptual distinctiveness of initial and final consonants in CVC syllables," J. Acoust. Soc. Am. **106**, 1555–1565.

Schroeder, M. R., Atal, B. S., and Hall, J. L. (**1979**). "Objective measure of certain speech signal degradations based on masking properties of human auditory perception," in *Frontiers of Speech Communication Research*, edited by B. Lindblom and S. Öhman (Academic, London), pp. 217–229.

Selkirk, E. O. (**1984**). *Phonology and Syntax* (MIT Press, Cambridge).

Simoncelli, E. P., and Olshausen, B. A. (**2001**). "Natural image statistics and neural representation," Annu. Rev. Neurosci. **24**, 1193–1215.

Sluijter, A. M. C., and van Heuven, V. J. (**1995**). "Effects of focus distribution, pitch accent and lexical stress on the temporal organization of syllables in Dutch," Phonetica **52**, 71–89.

Sluijter, A. M. C., and van Heuven, V. J. (**1996a**). "Spectral balance as an acoustic correlate of linguistic stress," J. Acoust. Soc. Am. **100**, 2471–2485.

Sluijter, A. M. C., and van Heuven, V. J. (**1996b**). "Acoustic correlates of linguistic stress and accent in Dutch and American English," 4th International Conference of Spoken Language Processing (University of Delaware), pp. 630–633.

Smiljanic, R. (**2004**). *Lexical, Pragmatic, and Positional Effects on Prosody in Two Dialects of Croatian and Serbian: An Acoustic Study* (Routledge, New York).

Strangert, E., and Heldner, M. (**1998**). "On the amount and domain of focal lengthening in Swedish," 5th International Conference on Spoken Language Processing (Sydney, Australia), pp. 3305–3308.

Syrdal, A. K., and Gopal, H. S. (**1986**). "A perceptual model of vowel recognition based on the auditory representation of American English vowels," J. Acoust. Soc. Am. **79**, 1086–1100.

't Hart, J., Collier, R., and Cohen, A. (**1990**). *A Perceptual Study of Intonation* (Cambridge University Press, Cambridge).

Tolhurst, G. C. (**1957**). "Effects of duration and articulation changes on intelligibility, word reception, and listener preference," J. Speech Hear Disord. **22**, 328.

Traunmüller, H., and Eriksson, A. (**2000**). "Acoustic effects of variation in vocal effort by men, women, and children," J. Acoust. Soc. Am. **107**, 3438–3451.

Turk, A. E., and Sawusch, J. R. (**1997**). "The domain of accentual lengthening in American English," J. Phonetics **25**, 25–41.

van Bergem, D. R. (**1993**). "Acoustic vowel reduction as a function of sentence accent, word stress, and word class," Speech Commun. **12**, 1–23.

van Son, R. J. J. H., and Pols, C. W. (**2003**). "Information structure and efficiency in speech production," Eurospeech/Interspeech 2003 (Geneva, Switzerland), pp. 211–214.

Watson, C. I., and Harrington, J. (**1999**). "Acoustic evidence for dynamic formant trajectories in Australian English vowels," J. Acoust. Soc. Am. **106**, 458–468.

Wright, R. (**2003**). "Factors of lexical competition in vowel articulation," in *Phonetic Interpretation: Papers in Laboratory Phonology VI*, edited by J. Local, R. Ogden, and R. Temple (Cambridge University Press, Cambridge), pp. 75–87.