Perception & Psychophysics
2002, 64 (1), 131-139

Perception of vocal effort and distance from the speaker on the basis of vowel utterances

ANDERS ERIKSSON and HARTMUT TRAUNMÜLLER
Stockholm University, Stockholm, Sweden

The sound pressure level of vowels reflects several nonlinguistic and linguistic factors: distance from the speaker, vocal effort, and vowel quality. Increased vocal effort also involves the emphasis of higher frequency components and increases in *F*0 and *F*1. This should allow listeners to distinguish it from decreased distance, which does not have these additional effects. It is shown that listeners succeed in doing so on the basis of single vowels if phonated, but not if whispered, and that they compensate for most of the between-vowel variation in level. The results obtained when listeners had to estimate vocal effort as well as distance suggest that an analysis of an utterance takes place at an early stage in auditory processing, before memories of episodes are stored.

In a series of experiments, Ladefoged and McKinney (1963) showed that listeners' judgments of the *loudness* of syllables were more closely correlated with the subglottal pressure with which they had been produced than with their sound pressure level (SPL). The variable that the subjects of Ladefoged and McKinney were estimating under the label of *loudness* was probably the effort with which the syllables had been produced. An increase in vocal effort involves an increase in subglottal pressure, by which the SPL and loudness of a speech signal increases and, normally, also its pitch. In addition to phonation, the articulation of speech is also affected by an increase in vocal effort, which results in additional acoustic variables being affected (Traunmüller & Eriksson, 2000). These have been shown to be important for the perception of vocal effort (Rundlöf, 1996; Traunmüller, 1997). Among them, the increases in the emphasis of high-frequency components, in fundamental frequency (*F*0), and in the frequency position of the first formant (*F*1) are especially important. As compared with intensity (or SPL), spectral balance (higher frequency emphasis) has also been shown to be a better correlate of linguistic stress (Sluijter, Shattuck-Hufnagel, Stevens, & van Heuven, 1995; Sluijter & van Heuven, 1996; Sluijter, van Heuven, & Pacilly, 1997). In a recent study, Yi, Kim, and Lee (2000) claimed that an utterance-initial rise and a final fall in *F*0 is a strong cue to an increase in vocal effort, but it may be the case that the effect is due to the increase in *F*0 alone.

For the (nonlinguistic) variable of listening distance, SPL decreases with increasing distance from the speaker, and in a free field, increases in distance have no additional effects. It is this kind of variation, and not variation in vocal effort, that can be mimicked by manipulating the SPL of a speech signal. SPL is closely related to a psychoacoustic variable for which the term *loudness* is well established, but loudness cannot, in general, be equated with (the inverse of) distance or with vocal effort.

A measure of vocal effort can be obtained by letting subjects rate the distance over which a speaker intends to communicate. In the experiments reported here, subjects were to estimate the communication distance between a speaker and the addressee and their own apparent distance from the speaker.

Previous research has shown that listeners are very accurate in distinguishing between these two types of variation in connected speech on the basis of sentence-length utterances. In an experiment by Wilkens and Bartel (1977), recorded sentences were presented via a loudspeaker in an anechoic chamber or via headphones. The sentences varied in the vocal effort with which they had been produced, the playback loudness level, and speaker familiarity. Subjects were to judge whether the playback level was too low or too high or correctly represented the original production. This task was preceded by a training period in which the subjects were to familiarize themselves with the task and to provide the experimenters with a baseline measure of the accuracy with which such matching can be performed. In a series of preliminary experiments, listeners had to compare the level of recorded sentences reproduced with the original SPL with copies that varied in SPL, in a pairwise comparison task, and say whether the comparison was weaker, was stronger, or matched the original. The subjects succeeded in detecting deviations from the original within less than 1 dB. They were more successful with familiar voices than with unfamiliar ones. Then the interstimulus interval was gradually increased to hours and then to days, and in the final experiment no reference at all was provided. Subjects nevertheless succeeded in recognizing the original levels within 2 dB. To be able to do this,

Correspondence concerning this article should be addressed to A. Eriksson, Department of Linguistics, Stockholm University, S-106 91 Stockholm, Sweden (e-mail: anders@ling.su.se).

the subjects must have been able to distinguish, quite accurately, between the vocal effort with which the sentences had been produced and the distance to the recorded speaker.

A more recent study by Rundlöf (1996, partly reported in Traunmüller, 1997) lends further support to this assumption. Stimuli used in the experiment were utterances recorded in an open field where a speaker communicated with a listener over distances varying between 0.3 and 187.5 m. A description of the procedures and an acoustic analysis of the effects of varying vocal effort may be found in Traunmüller and Eriksson (2000). These recordings were used in perception tests in which the task was to estimate the distance over which the speakers were perceived to communicate. Three series of stimuli were prepared, one series in which the original sound levels were used, one in which the levels were attenuated by 6 dB, and finally one in which all levels were equalized. The first two series were mixed in random order in Experiment 1, and the third series was used in Experiment 2. Two observations could be made. First, the distance estimates correlated well with the actual distances ($r = .918$ in Experiment 1, and $r = .914$ in Experiment 2). Second, the variation in level did not significantly influence the distance estimates. The two series used in the first experiment, differing in level by 6 dB, did not produce significantly different distance estimates. And the distance estimates obtained in Experiment 2, in which the level cue was eliminated altogether, did not differ significantly from those obtained in Experiment 1. The implication of these results is that listeners were perfectly able to separate variations in level from variations in effort although the perceived distance between the observer (the subjects in this experiment) and the speaker was not explicitly tested.

The question now arises, what is the minimal utterance on the basis of which a listener is able to estimate, with a reasonable degree of accuracy, the vocal effort of a speaker and the listener's own apparent distance to the speaker?

Utterances consisting of a single phonated vowel may already contain enough information to perform distance estimates. All the cues that have been shown to contribute significantly to the perception of vocal effort (Traunmüller, 1997)—namely, spectral emphasis, $F0$, and the formant frequencies (in particular, $F1$)—are present in single phonated vowels. Whispered vowels are deficient in these respects. They lack the spectral emphasis cue, as well as the $F0$ cue. They do contain formant frequency cues, but these are not reliable, since similar variations in formant frequencies can also result from small variations in pronunciation that are due to factors other than vocal effort. When presented with nothing more than a whispered vowel, subjects will have to base their judgments mainly on the level of the vowels, but in the absence of reliable additional cues, we can expect them to confuse the causes of variations in level.

For utterances consisting of a single vowel, there is, however, a third factor to be considered, often referred to as the intrinsic level of vowels. The SPL of different vowels produced with the same subglottal pressure is not the same. The reasons for this are well understood on the basis of the acoustic theory of speech production (Fant, 1960), which, in principle, allows us to calculate the level differences from the other characteristics of the vowels, albeit this is somewhat complex. Informal experiments in speech synthesis have repeatedly shown that listeners require the vowel-specific intensity variations to be reproduced in synthetic speech; otherwise, the impression is evoked that "somebody manipulates the volume control knob" while the speech is produced.

The observation by Ladefoged and McKinney (1963) that the *loudness* of syllables was more closely correlated with the subglottal pressure than with SPL also implies that the listeners compensated for most of the between-vowel variation in intrinsic level.

The fact that the SPL of a vowel is influenced, not only by vocal effort and listener distance (simulated by presentation level), but also by intrinsic level presents, however, a further complication. In addition to separating changes in SPL caused by vocal effort from changes caused by listening distance, subjects must also compensate for the variation caused by intrinsic level.

The questions posed in this study are, thus, the following. (1) Will subjects be able to distinguish between vocal effort and listening distance on the basis of isolated vowels? (2) Will they be able to ignore the variation in SPL caused by the intrinsic level variation?

The various aspects involved here may be understood within the framework of the modulation theory (Traunmüller, 1994, 2000). This theory is based on the tenet that when talking, speakers modulate their voice with speech gestures. Thus, the voice serves as a carrier signal, and the linguistic information is conveyed by its modulation. Subsequently, listeners are assumed to demodulate the acoustic signal in order to separate the linguistic, expressive, organic, and perspectival information in speech signals. The present experiments are primarily concerned with the perception of the expressive and perspectival qualities of vowels (vocal effort and distance), whereas linguistic and organic variation (intrinsic level and speaker) is merely a source of interference.

According to modulation theory, distance judgments are not based directly on the acoustic properties of the speech signal, but on those of an inferred carrier signal, which can be thought of as a neutral vowel whose properties are descriptive of the speaker's voice. In order for this to succeed, and to avoid interference from intrinsic level, it must be possible to infer the properties of the carrier signal with sufficient accuracy. Isolated whispered vowels do not contain enough information for such an inference. Some interference between intrinsic level and distance judgments is to be expected even when the vowels are phonated, since the carrier signal is not very accurately specified in the absence of a certain segmental variation.

In theoretical and experimental literature from the past decade, many researchers have taken exactly the opposite point of view and have claimed that the linguistic and non-

Table 1
Sound Pressure Levels (in Decibels) of the Chosen Tokens as a Function of Communication Distance, Speaker M.P. (Male)

	Phonated				Whispered			
	Distance (m)				Distance (m)			
Vowel	1.5	6	24	Mean	0.375	1.5	6	Mean
[i]	53.0	57.0	61.4	57.1	16.9	22.0	30.3	23.1
[y]	55.7	58.9	62.2	58.9	18.4	22.3	30.0	23.6
[u]	53.2	56.0	63.4	57.5	22.3	22.5	25.2	23.3
[æ]	55.8	59.9	64.3	60.0	27.2	31.4	28.0	28.8
[ɒ]	59.3	61.4	66.9	62.5	31.1	33.8	37.4	34.1
Mean	55.4	58.6	63.6	59.2	23.2	26.4	30.2	26.6

linguistic information that is conveyed by the speech signal are perceived and encoded in memory in an integral (not demodulated) fashion (Goldinger, 1996; Johnson, 1997; Pisoni, 1993). Even when this hypothesis has had to be rejected on the basis of experimental results, the claim has been maintained. Thus, Pisoni (1993, p. 118) observed that experimental results "demonstrate a form of implicit memory for a talker's voice that is distinct from the retention of the individual items used," but in the abstract of the same paper we nevertheless read that "taken together, the present set of findings are consistent with non-analytic accounts." In a more recent investigation, Nygaard, Sommers, and Pisoni (1995) failed to observe the strong effect of speaking rate that would be expected on this basis, and Bradlow, Nygaard, and Pisoni (1999) obtained evidence against other implications of such a point of view. All the evidence obtained in this line of research suggests that listeners separate the different kinds of information in speech signals at an early stage in processing in such a way as that described by the modulation theory. We shall return to this question in the General Discussion section of our results.

PREPARATORY EXPERIMENT

The aim of the experiment was to obtain a set of vowel sounds to be used in subsequent perception experiments. For this purpose, phonated and whispered versions of the Swedish names of the letters *i* [i], *ä* [æ], *a* [ɒ], *o* [u], and *y* [y], were produced at several levels of vocal effort. This was controlled by varying the communication distance between speaker and addressee. Subsequently, their sound pressure levels were measured in order to get hold of the between-vowel variation in levels that is due to intrinsic factors. This was achieved by comparing the actual levels of the vowels with the average of all the vowels produced by a given speaker in a given mode (phonated or whispered) at a given distance.

Method

Speakers. The vowels were produced by two adult speakers, one female (U.S.) and one male (M.P.). Both were teachers at the Department of Linguistics, Stockholm University.

Procedure. The vowel utterances and their variation in vocal effort were elicited by one of the experimenters asking the speakers from various distances for the name of a vowel letter he showed them (*i, ä, a, o, y*). The order of the letters was randomized for each distance, with the exception of a final dummy (*ö*), and each letter appeared twice. The distances were 1.5, 6, and 24 m for the phonated vowels and 0.375, 1.5, and 6 m for the whispered vowels. The shortest distances used (1.5 m for phonated vowels and 0.375 m for whispered vowels) was intended to represent normal effort. Subsequent distances were quadruples (powers of two) of the base distance. Although each vowel was produced twice under each condition (speaker, mode, distance, and phoneme), only one representative was used in the following perception experiments. The selection was based on criteria aimed at avoiding tokens with anomalies, such as partial voicing of the voiceless vowels, and the first of the vowels produced in a given condition was avoided.

Results and Discussion

The average levels, in decibels relative to an arbitrary reference, of the vowels per distance are listed in Tables 1 and 2 for the two speakers. These data include only those tokens that were used in the subsequent experiments, but the values are not very different from the averages obtained from all productions, reported previously by Eriksson and Traunmüller (1999). For voiced speech, the male speaker used a markedly smaller dynamic range than did the female speaker—8.0 dB, as compared with 13.7 dB, averaged over all vowels. This was mainly due to his relatively high SPL values at the shortest distance. In all other cases, the levels were lower in M.P.'s vowels than in U.S.'s. However, these differences cannot be ascribed to sex, since an investigation of the acoustic effects of variations in vocal effort (Traunmüller & Eriksson, 2000) showed the levels in the speech of adult male and female speakers to be very similar at all communication distances between 0.3 and 187 m.

In order to fully compensate for the acoustic effects of a free-field increase in communication distance by a factor of two, an increase in SPL by 6 dB would appear to be required. We can see that our speakers go only about halfway. However, as was mentioned in the introduction, a natural increase in vocal effort involves increases not only in SPL, but also in higher frequency emphasis, pitch, and $F1$ (Traunmüller & Eriksson, 2000). Since all these variations increase the audibility of the speech signal, speakers do not need to increase their SPL so much.

In Table 3, the mean SPL of each vowel (three tokens produced at different levels of vocal effort) is expressed in relation to the mean of the vowels of each speaker and mode

Table 2
Sound Pressure Levels (in Decibels) of the Chosen Tokens as a Function of Communication Distance, Speaker U.S. (Female)

	Phonated				Whispered			
	Distance (m)				Distance (m)			
Vowel	1.5	6	24	Mean	0.375	1.5	6	Mean
[i]	51.5	59.7	63.8	58.3	24.5	28.3	29.7	27.5
[y]	50.6	58.7	63.2	57.5	21.8	28.1	29.2	26.4
[u]	54.8	63.2	66.7	61.6	31.5	30.9	35.7	32.7
[æ]	52.5	59.5	68.6	60.2	37.2	33.1	35.3	35.2
[ɒ]	53.8	61.1	69.9	61.6	27.4	27.6	39.9	31.6
Mean	52.5	60.3	66.2	59.8	27.8	29.1	34.7	30.7

Table 3
Intrinsic Levels: Sound Pressure Levels Relative to the Mean for Each Speaker and Mode of Phonation

	Phonated		Whispered	
Vowel	M.P.	U.S.	M.P.	U.S.
[i]	−2.1	−1.5	−3.5	−3.2
[y]	−0.3	−2.3	−3.0	−4.3
[u]	−1.7	+1.8	−3.3	+2.0
[æ]	+0.8	+0.4	+2.2	+4.5
[ɒ]	+3.3	+1.8	+7.5	+0.9

of production. The table reveals substantial differences between the vowels and also between the two speakers.

Between-vowel differences of the kind observed here are to be expected on the basis of the acoustic theory of speech production (Fant, 1960). However, since only two speakers were used in the present experiment, the data do not allow any conclusions to be drawn as to whether the observed differences are due to sex or to other types of individual, between-speaker variations.

For the purpose of the following study it is, however, necessary neither for the speakers to be ideal representatives of male and female speakers in general nor for the vowels to be ideal representatives of their respective categories. The underlying assumption is that listeners should be able to reconstruct vocal effort and listening distance on the basis of the acoustic properties of the individual tokens alone, without having to make reference to internalized reference speakers or vowels.

EXPERIMENT 1

Method

Stimuli. For the perception experiments, one representative of each vowel, at each distance, in each mode, and by each speaker was used. Additional stimuli were obtained by modifying the SPLs of the phonated vowels by –6 and –12 dB and those of the whispered vowels by +6 and –6 dB. This was done in order to simulate variation in the subjects' distance from the speaker. This resulted in a total of 180 different stimuli (two speakers, two modes of production, five vowels, three levels of vocal effort, and three levels of presentation).

Listeners. Twenty-four paid listeners served as subjects. All of them were students at the linguistics department of Stockholm University, with Swedish as their first or, at least, their most often used language. Except for 1, no subject reported any known history of hearing disorders. The one who did was tested by standard audiometry, but no deviation from the normal threshold of hearing was found.

Procedure. There were two versions of the experiment. In the first version (Experiment 1A), 12 subjects were asked to estimate the distance over which the two participants in the exchange were communicating (communication distance). The subjects could only hear the speaker who pronounced the vowels. In the second version (Experiment 1B), another 12 subjects were asked to estimate their own apparent distance from the speaker (listening distance). This precluded the use of headphones, which tends to evoke the impression of the sound's coming from inside the head.

In both versions, the stimuli were presented to the listeners via a loudspeaker hung from the ceiling in one corner of an anechoic chamber. The subjects were seated in front of a computer in the opposite corner, 3.5 m away. The longest possible distance practically available was used in order to minimize the risk of producing distance cues connected with the distance to the loudspeaker. In order to avoid visual cues of this kind, the light in the chamber was dimmed.

Using a program designed for running perception experiments, each stimulus was presented once, in an order that was separately randomized for each listener. No feedback was given. Each run began with six stimuli presented for the subjects to acquaint themselves with the procedure.

Answers were to be chosen from a list of suggested distances ranging from 0.2 to 37 m for communication distance, and from 0.2 to 23 m for listening distance. The ranges were divided into roughly equal steps on a log scale, with 32 and 29 values, respectively. The distances used in the communication distance task represented the actual distances used when the vowel utterances were recorded but extended in both ends in order to allow for a reasonable degree of over- and underestimation. The range of distances used for the listening distance estimates would have had to cover roughly a factor of 16 to correspond exactly to the presentation level manipulations. Here, a slightly wider margin at both ends was chosen because of our more limited experience with this kind of task, to ensure that the range would not restrict subjects' decisions. Pretest trials using the authors as subjects suggested that the distances were appropriate for the tasks.

Results

The results obtained from each listener were subjected to individual analysis. This revealed substantial between-listener variation. With the phonated stimuli, a few listeners showed no significant positive correlation between original level and communication distance (2 of 12 listeners in Experiment 1A) or no significant negative correlation between presentation level (amplification) and listening distance (3 of 12 listeners in Experiment 1B). The results obtained from these subjects were excluded from further consideration. Our interpretation of their results was that these subjects had simply not understood the task. (An analysis including these data showed that the only effect was adding more noise, whereas the general results remained the same.)

The following analysis considers only the median values obtained from the responses of the remaining subjects for each stimulus. The distance ratings in meters were converted to base 2 logarithms (henceforth, 2-logarithms). These values were then compared with (1) the 2-logarithm of the original sound pressure and (2) the 2-logarithm of the level modification. We chose 2-logarithms as a common scale in order to facilitate comparisons between different factors.

The medians of the ratings obtained in Experiment 1A, in which the listeners judged the communication distance, are shown in Figure 1, where they are plotted against Variables 1 and 2. The medians of the ratings obtained in Experiment 1B, in which the listeners judged the listening distance, are plotted against the same variables.

In order to gain some insight into the possible effects of differences in intrinsic level, Variable 1 was split up into two parts: (1a) a basic part that can be assumed to reflect the speaker's vocal effort and was calculated as the average level of the vowels produced by a given speaker in a given mode at a given distance and (1b) a supplementary part that reflects all between-vowel variation for a given speaker, mode, and distance. The values chosen were those that resulted from the preparatory experiment (Tables 1 and 2).

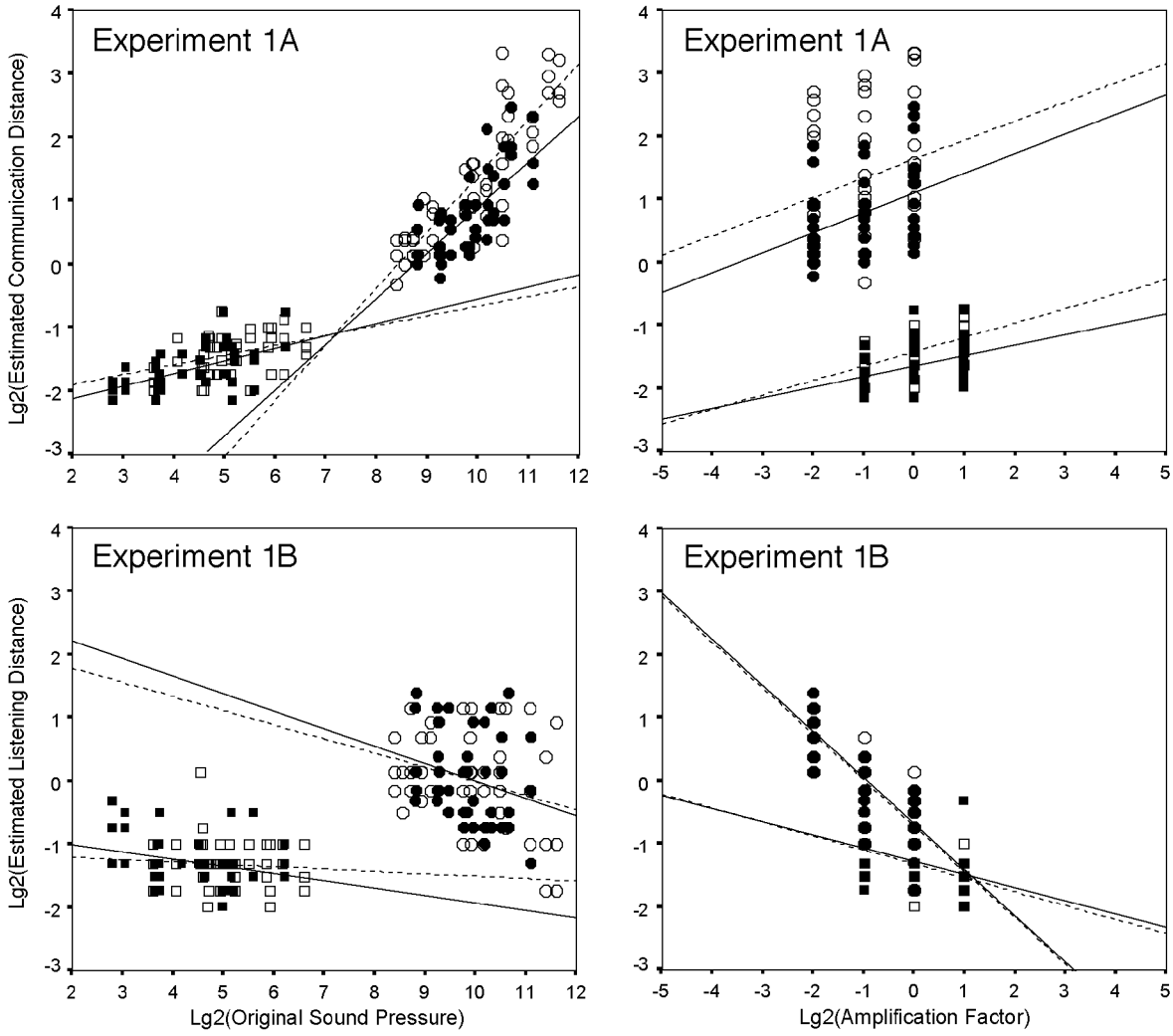

Figure 1. Average distance rating for each stimulus plotted against original sound pressure level (left panels) and amplification (right panels) shown for Experiment 1A (communication distance, upper panels) and Experiment 1B (listening distance, lower panels). Circles, phonated speech; squares, whispering. Regression lines fitted to the results obtained with each speaker. Filled symbols, solid lines, speaker M.P. (male), unfilled symbols, dashed lines, speaker U.S. (female).

Table 4 summarizes the results of regression analyses in which the 2-logarithm of the distance rating was taken as the dependent variable and 1a, 1b, and 2 as the independent variables. The analysis was performed separately for each speaker and mode of phonation. The values entered into the table show the perceptual effect of a 6-dB increase in the level of each independent variable. This is expressed in powers of 2. Thus, a value of +1.0 would mean that a 6-dB increase in SPL would cause the distance estimate in meters to double.

Discussion

If vocal effort were simply a matter of SPL, an increase of 6 dB would be required to fully compensate for a doubling in communication distance. However, as was mentioned in the introduction, an increase in vocal effort is accompanied by changes in a variety of parameters that all contribute to an emphasis of the perceptually more important frequency components above the fundamental of the speech signal. Therefore, SPL does not need to be increased by a full 6 dB. In the experiments described in Traunmüller and Eriksson (2000), it was found that an increase of SPL for the voiced segments of an utterance by 4.6 dB was required in order for listeners to perceive a distance doubling. The value obtained in Experiment 1A with phonated vowels was larger, 6.6 dB (6/0.903).

In a free field, a doubling in listening distance results in an SPL decrease by 6 dB. The results of Experiment 1B show that listeners require an attenuation of 8.2 dB (6/0.730) in order to double the estimate of their own distance from the speaker with phonated vowels.

These differences (6.6 vs. 4.6 and 8.2 vs. 6.0) may be explained by the considerable increase in difficulty of the task when judgments have to be based on only a minimal

Table 4
Summary of Regression Analyses Showing Weights and Significance Levels in Experiment 1

	Phonated				Whispered			
	Communication Distance $r^2 = .80$		Speaker Distance $r^2 = .74$		Communication Distance $r^2 = .59$		Speaker Distance $r^2 = .30$	
	Weight	p	Weight	p	Weight	p	Weight	p
Effort +6 dB	**+0.903**	<.001	−0.227	<.001	**+0.320**	<.001	−0.249	<.01
Intrinsic SPL +6 dB	+0.467	<.01	−0.308	<.05	+0.131	<.001	−0.026	n.s.
Amplification +6 dB	+0.308	<.001	**−0.730**	<.001	+0.198	<.001	**−0.214**	<.001
Speaker		<.001		n.s.		n.s.		n.s.

Note—The target variable is in boldface.

utterance, one isolated vowel. In this case, there is a greater interference by the other two variables, intrinsic level and either vocal effort or presentation level (amplification).

Ideally performing listeners would attach a high weight to vocal effort in Experiment 1A (ideally, >1.0) and to amplification in Experiment 1B (ideally, 1.0). They would attach a weight of 0.00 to amplification in Experiment 1A and to vocal effort in Experiment 1B. The results obtained with phonated stimuli showed that the weight listeners attached to the interfering variables, amplification (in Experiment 1A) and vocal effort (in Experiment 1B) was about a third of that attached to the target variable. The weight of intrinsic level was about half that of the target variable. With the whispered stimuli, performance was, as expected, less "ideal." The weight of the target variable was low, and there was relatively more interference from amplification in Experiment 1A and from effort in Experiment 1B, but the interference from intrinsic level was weaker than that observed with the phonated vowels. It was especially low in Experiment 1B, in which the subjects attached about the same weight to the other interfering cue (effort) as to the target variable (amplification). The contribution of the intrinsic cue was significantly different from 0 in three of the four partitions. We have to conclude that the listeners were not completely successful in compensating for the intrinsic level variations but that they compensated for the larger part of them.

There were some differences between the two speakers not only in the acoustic data of their vowels, but also in the distance estimates by the listeners. These were significant only for the phonated vowels in Experiment 1A.

EXPERIMENT 2

The results obtained in Experiment 1 indicated that although the listeners were able to distinguish between cues for vocal effort and their own apparent distance from the speaker (simulated as presentation level variation), there was considerable confusion between the two. In each of the two versions of Experiment 1, only one of the questions (communication distance or listening distance) was asked. This may have caused the listeners to be insufficiently aware of the distinction even though they might, in principle, have been able to keep the two types of distances apart better than was indicated by the results. The second experiment was constructed to throw some light on this question. Can the results be improved by simply making subjects more aware of the two dimensions involved? To this end, the design was changed so that the listeners had to judge both distances for each stimulus.

This involves an increase in memory load that may be interesting in itself. Listeners have to keep their impression of the stimulus in memory while answering the first of the two questions, and they have to base the second answer on the picture of the stimulus in their memory. Therefore, the results may tell us something about how much detailed information about the stimuli is retained in memory. In order to explore these aspects, two versions of the experiment were constructed, one in which the vocal effort estimate was to be made first and one in which the questions were presented in the reversed order.

Method

Stimuli. The stimuli were identical with those used in Experiment 1.

Listeners. Forty paid listeners served as subjects. All of them were students at the linguistics department of Stockholm University, with Swedish as their first or, at least, their most often used language.

Procedure. The same stimuli as those used in Experiment 1 were used in four versions of an additional perception experiment. Two versions contained only the stimuli produced by the female speaker, and two others those of the male speaker. The method of stimulus presentation and response collection was the same as that in Experiment 1.

The new experiment differed also from the previous one in that the subjects had to estimate the communication distance as well as their own apparent distance from the speaker for each presentation of a vowel. For each speaker, one group of 10 listeners had to estimate the communication distance before their own distance from the speaker, and the remaining 10 listeners had to estimate the distances in the opposite order.

The reason for dividing up what could have been one set of stimuli containing both the male and the female stimuli into two sets was to minimize the effects of fatigue on the part of the listeners. It was estimated that a high degree of concentration by the listeners was necessary to solve the task successfully and that the advantage of making the listeners aware of the fact that two judgments were involved might be weakened or canceled out by fatigue in a long and tiring session.

The two lists of suggested distance values were the same as those in the previous experiments.

Table 5
Summary of Regression Analyses Showing Weights and Significance Levels in Experiment 2

	Phonated				Whispered			
	Question 1, Communication Distance $r^2 = .85$		Question 2, Speaker Distance $r^2 = .70$		Question 1, Communication Distance $r^2 = .70$		Question 2, Speaker Distance $r^2 = .25$	
	Weight	p	Weight	p	Weight	p	Weight	p
Effort +6 dB	**+0.988**	<.001	−0.227	<.001	**+0.536**	<.001	−0.072	n.s.
Intrinsic SPL +6 dB	+0.221	<.05	−0.092	n.s.	+0.194	<.001	−0.022	n.s.
Amplification +6 dB	+0.271	<.001	**−0.598**	<.001	+0.472	<.001	**−0.204**	<.001
Speaker		<.001		<.05		n.s.		n.s.

	Phonated				Whispered			
	Question 1, Communication Distance $r^2 = .85$		Question 2, Speaker Distance $r^2 = .80$		Question 1, Communication Distance $r^2 = .65$		Question 2, Speaker Distance $r^2 = .44$	
	Weight	p	Weight	p	Weight	p	Weight	p
Effort +6 dB	**+0.847**	<.001	+0.009	n.s.	**+0.527**	<.001	−0.284	<.01
Intrinsic SPL +6 dB	+0.133	n.s.	−0.288	<.05	+0.054	n.s.	−0.124	<.05
Amplification +6 dB	+0.164	<.001	**−1.011**	<.001	+0.397	<.001	**−0.364**	<.001
Speaker		<.01		n.s.		n.s.		n.s.

Note—The target variable is in boldface.

Results

In the version using the male speaker with listening distance as the first question, the responses obtained from 3 subjects were excluded since they showed no significant positive correlation with the target variable. There were no exclusions in the three other versions.

The medians of the responses by the 10 subjects in each group were calculated for each stimulus, after conversion of the distance estimates from meters to 2-logarithms. As in Experiment 1, these values were then compared with the 2-logarithms of the original sound pressure and the amplification.

If plotted in the same way as in Figure 1 for Experiment 1, the results of Experiment 2 look very similar. We will, therefore, not present the results of this experiment in the form of a diagram. All the information that is essential for a comparison of the results obtained in the four versions of this experiment with those obtained in Experiment 1 can be found in Tables 5A and 5B, which are analogous to Table 4. They show the perceptual effect of a 6-dB increase in the level of each independent variable.

One may also consider how much of the variance is explained by the underlying variables. In Figure 2, the variance explained by intrinsic level is shown for both experiments. It is of particular interest that the variance explained by this factor approaches zero in the answers to the second question in Experiment 2.

Discussion

The results from this experiment may be looked upon from two different points of view. (1) Will subjects perform better when made aware that two different factors, communication distance and listening distance, are involved? (2) Will subjects perform better on the first question than on the second, where keeping the impression of the sound in memory is involved?

It may be observed that the performance was indeed somewhat better for the first question, both in terms of explained variance and in terms of the weight attached to the target variable. With the phonated vowels, the increases in level necessary for a doubling of the communication dis-

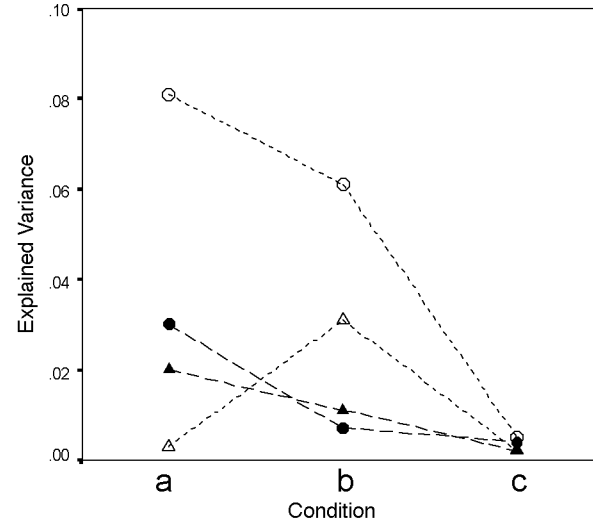

Figure 2. Variance explained by the intrinsic level variable in Experiment 1 (a), and Experiment 2, first question (b) and second question (c). Circles, communication distance estimate; triangles, listening distance; filled symbols, phonated vowels; open symbols, whispered vowels.

Table 6
Weight of the Cues for the Target Variables (Boldface in Table 5) in the Answers to the Second Question Expressed as a Percentage of the Weight of the Same Cues When the Question was Asked First

Cue	Phonated (%)	Whispered (%)
Effort as a cue to communication distance	86	98
Amplification as a cue to listening distance	59	56

tance estimate were 6.1 and 7.1 dB (as compared with 6.6 in Experiment 1 and 4.6 with sentences), and the corresponding values for a doubling of the listening distance estimate were −5.9 and −10.0 dB (as compared with −8.2 in Experiment 1 and −6.0 as a theoretical ideal).

Although the overall performance became worse, there was less interference in the responses to the second question, as compared with the first—in particular, interference from intrinsic level. This can be seen in Table 5 and even more clearly in the r^2 values shown in Figure 2, which were negligibly small (<.005) for intrinsic level in all four cases of second questions. A possible explanation for this observation will be offered in the General Discussion section.

GENERAL DISCUSSION

One of the questions that motivated this study was whether listeners are able to distinguish between two types of variation in speech, variation in vocal effort and variation in their own apparent distance from a speaker. The results in Experiment 1 showed that listeners are indeed able to do this. They also showed, however, that performance was far from perfect, owing to interference between the various factors involved. It was thought that one possible explanation for the interference effects was that the listeners were insufficiently aware of the fact that two types of distances were involved. In the second experiment, this was remedied by explicitly asking both questions, the expectation being that this would enhance performance, at least for the first question asked. As was shown above, this expectation was met.

Comparing the weights and r^2 values obtained in the responses to the first question with those from the second question and with those from Experiment 1, we can see the following.

1. With phonated vowels, variation in vocal effort was interpreted mostly as a variation in communication distance. To some extent, an increase in vocal effort was misinterpreted as a decrease in listening distance, as could be expected from the increase in SPL. However, in some exceptional instances, subjects mistook an increased communication distance for an *increased* listening distance. Perhaps, in these cases, subjects mistakenly thought of the speaker as speaking to *them.*

2. With phonated vowels, variation in presentation level was interpreted mainly as a variation in listening distance. However, in all conditions, variation in SPL owing to variation in presentation level and intrinsic level was to some extent misinterpreted as a variation in vocal effort. With whispered vowels, the listeners appear to have ascribed a variation in SPL to each of the two possible causes roughly to the same extent.

3. When the subjects had to evaluate both vocal effort (communication distance) and amplitude (listening distance) of the same stimuli, the estimation of vocal effort was not much affected by the subjects' prior task of estimating the amplitude of the stimuli, whereas the prior task of estimating vocal effort resulted in a clearly poorer performance in subsequent amplitude estimations (see Table 6).

This result shows that the amplitude of utterances is less persistently retained in memory than is their vocal effort. Such a differential result appears to be incompatible with models of episodic memory that assume utterances to be retained without prior analysis into their perspectival, organic, expressive, and linguistic components. An analogous result, also incompatible with such models of episodic memory, was obtained by Bradlow et al. (1999), who observed that listeners' performance in deciding whether a word had been presented previously in a list of words was not significantly affected by variations in amplitude (perspectival), whereas listeners were less accurate when there were variations in speech rate (expressive) and in speaker (organic variation).

When we consider what tends to be most important for a listener to perceive, it makes sense that perspectival variation does not interfere so much and that the retention of the perspectival aspects is less persistent than that of the communicative aspects of an utterance.

4. With both phonated and whispered vowels, intrinsic level interfered to a similar extent with judgments of effort and listening distance. This interference was distinctly less in the answers to the second question, as compared with those to the first and when only one question was asked. A model of episodic memory that assumes utterances to be retained in memory without discrimination of its different aspects does not offer an explanation for such a reduction in interference.

The phenomenon can be understood only if we allow for differential decay of memory traces reflecting different aspects of utterances. When the listeners answered the first question in Experiment 2, they can be assumed still to have had access to a very accurate and not yet analyzed sensory representation of the sound. The perception of the direction from which a sound reaches a listener requires such an accurate representation. However, this also appears to be the source of the interference, by the intrinsic between-vowel variation in SPL, observed in the answers to the first question.

When the subjects answered the second question, most of the detailed information appears already to have faded away, so that the amount of interference from intrinsic between-vowel variation was reduced substantially, although the overall performance became slightly worse.

The phenomena observed in these experiments can be understood if it is assumed that an analysis into linguistic, expressive, organic, and perspectival components, such

as the modulation theory (Traunmüller, 1994) attempts to describe, takes place at an early stage in auditory processing, before memories of episodes are stored. The present experiments have shown this in the pattern of interference obtained when subjects had to evaluate an expressive and a perspectival variable.

REFERENCES

Bradlow, A. B., Nygaard, L. C., & Pisoni, D. B. (1999). Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Perception & Psychophysics,* **61**, 206-219.

Eriksson, A., & Traunmüller, H. (1999). Perception of vocal effort and speaker distance on the basis of vowel utterances. In J. J. Ohala, Y. Hashegawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.) *Proceedings of the XIVth International Congress of Phonetic Sciences* (Vol. 3, pp. 2469-2472).

Fant, G. (1960). *Acoustic theory of speech production* (Vol. II in *Description and analysis of contemporary standard Russian*, R. Jakobson & C. H. van Schooneveld, eds.). The Hague: Mouton.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition,* **22**, 1166-1183.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 145-165). New York: Academic Press.

Ladefoged, P., & McKinney, N. (1963). Loudness, sound pressure and sub-glottal pressure in speech. *Journal of the Acoustical Society of America,* **35**, 454-460.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1995). Effects of stimulus variability on perception and representation of spoken words in memory. *Perception & Psychophysics,* **57**, 989-1001.

Pisoni, D. B. (1993). Long-term memory in speech perception: Some new findings on talker variability, speaking rate and perceptual learning. *Speech Communication,* **13**, 109-125.

Rundlöf, J. (1996). *Perceptuella ledtrådar vid auditiv bedömning av avståndet mellan talare och lyssnare* [Perceptual cues in auditory judgments of the distance between speaker and listener]. Unpublished master's thesis, Stockholm University, Stockholm.

Sluijter, A. M. C., Shattuck-Hufnagel, S., Stevens, K. N., & van Heuven, V. J. (1995). Supralaryngeal resonance and glottal pulse shape as correlates of stress and accent in English. In K. Elenius & P. Branderud (Eds.), *Proceedings of the XIIIth International Congress of Phonetic Sciences* (Vol. 2, pp. 630-633). Stockholm: Kungliga Tekniska Högskolan and Stockholm University.

Sluijter, A. M. C., & van Heuven, V. J. (1996). Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America,* **100**, 2471-2485.

Sluijter, A. M. C., van Heuven, V. J., & Pacilly, J. J. (1997). Spectral balance as a cue in the perception of linguistic stress. *Journal of the Acoustical Society of America,* **101**, 503-513.

Traunmüller, H. (1994). Conventional, biological and environmental factors in speech communication: A modulation theory. *Phonetica,* **51**, 170-183.

Traunmüller, H. (1997). Perception of speaker sex, age, and vocal effort. *Phonum* (Vol. 4, pp. 183-186). Umeå: Umeå University, Department of Phonetics.

Traunmüller, H. (2000). Evidence for demodulation in speech perception. In *Proceedings of the 6th International Conference on Spoken Language Processing* (Vol. III, pp. 790-793). Beijing: Chinese Academy of Sciences, Institute of Acoustics.

Traunmüller, H., & Eriksson, A. (2000). Acoustic effects of variation in vocal effort by men, women, and children. *Journal of the Acoustical Society of America,* **107**, 3438-3451.

Wilkens, H., & Bartel, H.-H. (1977). Wiedererkennbarkeit der Originallautstärke eines Sprechers bei elektroakustischer Wiedergabe [Perceptibility of the original loudness of a speaker through electroacoustic playback]. *Acustica,* **37**, 45-49.

Yi, S., Kim, H. S., & Lee, O. G. (2000). Glottal parameters contributing to the perception of loud voices. In *Proceedings of the 6th International Conference on Spoken Language Processing* (Vol. IV, pp. 580-583). Beijing: Chinese Academy of Sciences, Institute of Acoustics.

(Manuscript received July 19, 2000;
revision accepted for publication March 21, 2001.)