International Journal of Occupational Safety and Ergonomics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tose20

Effects of Synthesized Voice Warning Parameters on Perceived Urgency

Kyung S. Park[a] & Phil S. Jang[a]
[a] Korea Advanced Institute of Science and Technology, Korea
Published online: 08 Jan 2015.

INTERNATIONAL JOURNAL OF
OCCUPATIONAL SAFETY AND ERGONOMICS 1999, VOL. 5, NO. 1, 73–95

Effects of Synthesized Voice Warning Parameters on Perceived Urgency

Kyung S. Park
Phil S. Jang

Korea Advanced Institute of Science and Technology, Korea

The effects of synthesized voice warning parameters on perceived urgency were examined in order to build a detailed and usable description of the relation between the parameters of synthesized voice warnings and perceived urgency. Ten native and 10 non-native English speakers participated in 4 experiments to evaluate and quantify the effects of the voice parameters. The results showed that speech rate, average fundamental frequency (F_0), voice type, and fundamental frequency contour have clear effects on the perceived urgency of synthesized voice warnings. The effects of quantitative parameters on perceived urgency were scaled using an application of Stevens's power law (1957). In addition, the results showed significant differences in the perceived urgency of average F_0 and F_0 contour types between native and non-native English speakers. Implications of the results for the design and improvement of synthesized voice warnings are discussed.

warnings synthesized voice perceived urgency magnitude estimation
Stevens's power law

1. INTRODUCTION

For the last few decades, speech synthesis has been used successfully in a number of important applications such as the generation of warning messages in high workload environments such as aircraft cockpits. Recent advances in speech synthesis have made it possible to reproduce human speech at relatively low cost, and commercial speech synthesizers provide control by the user or the systems designer over the voice parameters such as fundamental frequency (pitch) and speech rate. This

Correspondence and requests for reprints should be sent to Phil S. Jang, Department of Industrial Engineering, Korea Advanced Institute of Science and Technology, 373-1 Kusong-dong, Yusong-gu, Taejon 305–701, Korea. E-mail: <roxe@chollian. net>.

availability has generated human factors considerations in terms of performance with, preference for, and appropriate use of synthesized voice warnings. Changes in the parameters of the synthesized voice could be expected to have effects not only on speech intelligibility, but also on the psychological appropriateness of speech such as its perceived urgency. There has been much research into the effects of emotion on the acoustic characteristics of human speech and the relations between perceived urgency and various types of warnings such as non-verbal auditory warning signals and warning labels. However, there have been no research studies about the perceived urgency of synthesized voice warnings. In this paper, we examine the relations between the various operational parameters of synthesized voice warnings and the perceived urgency of speech. This issue is important and the results would be useful for designing new warning signals and improving existing warnings of industrial, military, and public applications.

1.1. Synthesized Voice Warnings

Literature suggests that there are potential advantages of a proper use of a well-designed voice warning system. As Barber, Stanton, and Stockley (1992) pointed out, the advantage of speech displays, over other forms of displays, comes from their "public" nature and information-bearing properties. The public nature of speech displays is illustrated by its omnidirectionality; like other auditory displays, speech can be heard by everybody in a particular area and does not require a fixed display point or a person's orientation. Further, speech can be employed in situations that prohibit other types of displays such as low lightning levels. These points relate to the "eyes free, hands free" potential of speech and other auditory warning signals.

Speech warnings can function not only as signals but also as information carriers (Barber et al., 1992). They allow information to be presented to people in a familiar manner. This means that, unlike tonal displays, spoken messages do not require extensive learning. In addition, research has shown that a synthesized voice warning is inherently alerting (Byblow & Corlett, 1989). Thus, a synthesized voice warning can alert and inform in the same action. Edworthy (1994) argued that there are ways of making non-verbal warnings more sophisticated, so that they inform as well as alert. However, their meanings still have to be learned and trained beforehand. Speech has also been argued to be of use in situations of high workload such as aircraft cockpits (Simpson, McCauley, Roland, Ruth & Williges, 1987). In fighter aircraft flying at low levels, pilots have to maintain visual surveillance of their environment and they may not be able to attend to all the information on the visual displays in their cockpits. Speech can unload the pilot's information processing channels by employing an additional channel of communication (Aretz, 1983).

Whereas synthesized voice warnings have potential advantages over other types of warnings, the effectiveness of synthesized voice warning has not been clearly proved. As Edworthy (1994) pointed out, there are intrinsic advantages in using auditory (both verbal and non-verbal) warnings over written warnings. Otsubo (1988) showed that written warnings are seen and read, but are not necessarily complied with. In comparison with synthesized voice and other types of auditory warnings, earlier research has showed contradictory results. Byblow and Corlett (1989) found that if a synthesized voice warning begins with a redundant word, participants respond more quickly than if it begins with an alerting tone. There are studies that showed the alerting tone simply increases response time, and is therefore to all intents and purposes redundant (Hakkinen & Williges, 1984; Simpson & Williams, 1980). However, some experimental studies showed that, with better design and training, the use of attention getting sounds (alerting tone) can actually reduce response time and aid performance (James & James, 1989; Wheale, 1983).

Irving (1981) found that pilots responded more quickly and accurately to voice warnings than to tone warnings. However, Wheale (1983) found that synthesized voice warnings produced slower reaction times than other types of auditory warnings. Wheale's findings have resulted in the continued use of tone warnings and the continued research and refinement of synthesized voice warnings.

Most commercial speech synthesizers provide for the control of average fundamental frequency and speech rate by the user and the systems designer. Some speech synthesizers even provide for the control of formant frequencies, fundamental frequency contour, and laryngealization. Speech rate and fundamental frequency contour have been found to affect the intelligibility for a human receiver (Larkey & Danly, 1983; Simpson & Marchionda-Frost, 1984). In addition, Simpson and Marchionda-Frost (1984) argued that there is an alerting effect of higher voice pitch (fundamental frequency) in their study about the rate of speech and the effects of fundamental frequency on intelligibility. In natural speech, raised fundamental frequency, accompanied by changes of timing and increased amplitude, is used to signal emphasis and to alert. Studies of the effects of emotion on the acoustic characteristics of human speech have shown that fundamental frequency contours, average values, and ranges of fundamental frequency differ from one emotion to another (Streeter, Macdonald, Apple, Krauss, & Galotti, 1983; Williams & Stevens, 1972). These findings imply that synthesized voice parameters could be expected to have effect not only on intelligibility, but also on psychological appropriateness such as the perceived urgency of speech. Further, these findings show the possibilities of urgency mapping on synthesized voice warnings.

1.2. Urgency Mapping

The most urgent situations could be signaled by the most urgent warning sounds and the least urgent situations by the least urgent warning sounds. In order to do this, knowledge of the effects of individual sound parameters on perceived urgency is required. The experiments described in this study were to determine the effects of synthesized voice warning parameters on perceived urgency in order to build a detailed and usable description of the relation between synthesized voice warnings and perceived urgency. In Experiment 1, the parameters influencing perceived urgency were identified. These include speech rate, average fundamental frequency (F_0), voice type, and fundamental frequency contour. In Experiments 2 and 3, the relation between quantifiable parameters and perceived urgency were quantified using the application of Stevens's power law (1957). This power law is stated as

$$S = kO^m,$$

where S is sensation magnitude, O is stimulus intensity, k is an arbitrary constant determining the scale unit, and m is the power exponent, which depends on the sensory modality and stimulus conditions. The value of the exponent m indicates the power of the relationship between the subjective and objective values. The higher the value of m, the greater the change in the subjective judgment of urgency for a specific change in the relevant voice parameters. This value is typically less than 1 and as small as 0.33 for brightness (Stevens & Stevens, 1963) and loudness (Stevens, 1955), and as large as 3.5 for electric shock on the fingertip

(Stevens, Carton, & Shickman, 1958). These show that the sensation of electric current through the fingertips increases very rapidly according to stimulus intensity, whereas brightness grows very slowly as stimulus energy increases. The final experiment, Experiment 4, was concerned with the relation between F_0 contour and perceived urgency.

1.3. Magnitude Estimation

In all experiments conducted in this study, magnitude estimation was used to measure the perceived urgency of the participants. The method of psychophysical ratio scaling most frequently used in current investigations is the method of magnitude estimation. It has been used to describe the relationship between sensation magnitude and stimulus magnitude, such as brightness (Stevens & Stevens, 1963), loudness (Stevens, 1955), and electric shock on the fingertip (Stevens et al., 1958). Recent research has demonstrated that an application of this method is a viable experimental technique to quantify the relations between perceived urgency and non-verbal auditory warning parameters (Edworthy, Loxley, & Dennis, 1991; Hellier, Edworthy, & Dennis, 1993). It produces proportional, ratio preserving measures of opinion strength, thereby providing researchers with legitimate access to powerful statistical tools such as regression analysis for testing quantitative hypotheses. The participant in a magnitude estimation task is required to match numbers (numerical estimation), line lengths (line production), or forces of handgrip in proportion to the intensities of stimulus. Stevens (1958) described two main ways of applying the magnitude estimation technique to a scaling problem. In one, if the numerical estimation is used, the participant is presented with a standard stimulus and told that the sensation it produces has a certain numerical value (modulus), such as 10. In subsequent trials, other stimuli are presented and the participant assigns numbers to his or her sensations relative to the value of the modulus. The participant is instructed to make his or her judgements reflect how many times greater one sensation is than another (the ratio between the two sensations).

In the other version of the method, the modulus is not defined by the experimenter. The stimuli are randomly presented to the participant, who assigns numbers to the sensations in proportion to their magnitudes. As the participant is readily able to establish his or her own modulus, psychophysical scales with or without an experimenter-defined modulus are in close agreement (Gescheider, 1985). In this study, magnitude estimation with standard (experimenter-defined) modulus was used. In the following experiments, two techniques—numerical estimation (*NE*) and line production (*LP*, drawing lines to express the strength of one's impressions)—were used as quantitative response measures for scaling the perceived magnitude of stimuli. That is, the participant was required to draw a line and match a number in proportion to the perceived urgency of a warning signal. Numerical estimation and line production are widely used response measures in both sensory and social psychological scaling. As Stevens and Guirao (1963) pointed out, *NE* and *LP* are well-documented psychological modalities that are linearly related. Using these two methods of magnitude estimation rather than one, the cross-modality matching paradigm developed by Stevens (1966) can be used. The cross-modality matching paradigm provides an elegant method to confirm the power law and to verify the characteristic exponent relating stimulus magnitude to the magnitude of subjective response.

2. PRETEST AND PARTICIPANT SELECTION

Throughout this study, the magnitude estimation technique was used to record perceived urgency estimates made by the participants. This technique is based on the ability of a participant to make proportional judgements of stimuli magnitudes. To familiarize the participants with this technique and to validate their ability to make proportional judgements, a pretest was conducted prior to the main experiments. In addition, the results of the pretest were used to calibrate and correct the results of the main experiments for regression bias.

2.1. Method

2.1.1. Participants

A total of 24 undergraduate and graduate students participated in the pretest. Eleven of them were native English speakers, the others were Koreans who had studied English as a foreign language for over 10 years. There were 20 male and 4 female participants between the ages of 19 and 31.

2.1.2. Procedure

To scale the perceived urgency of the stimuli in the main experiments, participants used two response modalities: numerical estimation and line production. In the pretest, the participants' abilities to use both modalities were assessed. First, each participant was asked to read instructions adapted from Lodge (1981). They were then given sheets of paper that contained lines of various lengths. One of the lines was the reference line and it was given a numeric reference value. Participants were asked to estimate how much longer or shorter the other lines are in comparison with the reference line by giving each line a number compared to the reference value. Next, the participants were asked to draw lines without a ruler corresponding to the various numbers listed in random order.

2.2. Results

It is well established that the two response modalities, line production and numeric estimation, are linearly related. If the participant understands and uses the two response measures properly to make proportional judgement, the psychological functions derived from the line production of numeric stimuli and the numeric estimates of line length stimuli should be linearly related. The relations between the stimuli and the responses for individual participants were evaluated by regressing them in log-log coordinates. Next, the slopes were tested whether or not they fall within the 95% confidence limits constructed around the theoretical value of 1.0. Of the 24 participants, 20 participants successfully passed the pretest, and the 4 participants who failed to make related proportional judgements were excluded from the following experiments.

3. EXPERIMENT 1

The main objective of this experiment was to identify the parameters of synthesized voice that have influence on the perceived urgency of verbal warnings. The investigated parameters were speech rate, average fundamental frequency (F_0), voice type (male-sounding and female-sounding voice), and fundamental frequency contour. An important factor that was not investigated in this study was the overall intensity of the synthesized voice warnings. This factor was kept constant for each stimulus because it is related to the intelligibility of the stimuli and dependent on the background noise of the situation being signaled. In that situation, the appropriate intensity for a warning signal would be kept constant and thus the changes of perceived urgency should be achieved by some means other than changes of intensity.

3.1. Method

3.1.1. Participants

Twenty participants who had successfully completed the pretest participated in the experiments. Prior to the experiment, all participants were asked whether they had any hearing difficulties. No further audiometric tests were performed, because they were believed to be unnecessary, in that the stimuli were presented at a constant physical intensity level. The only likely effect of a participant with some hearing loss would be a depression of the mean of the magnitude estimation across all of that participant's responses.

Ten of the participants were native speakers of English and the other 10 were Koreans who had studied English as a foreign language for over 10 years. There were 17 male and 3 female participants between the ages of 19 and 29 (M = 24.75, SD = 2.5).

3.1.2. Apparatus

The synthesized voice warnings used in each experiment were generated by the TextAssist (Creative Labs) text-to-speech synthesizer. The synthesizer was driven by a Pentium processor-based personal computer equipped with a Sound Blaster (AWE32) sound board. The synthesized voice warnings were prerecorded on the same kind of PC at a sampling rate of 20 kHz and presented to participants through Phillips SBC3375 earphones (circumaural type).

Participants were given typewritten instructions, response sheets, and a pencil. The response sheet allowed a maximum line length of 360 mm.

3.1.3. Stimuli

Voice warnings should be worded as short phrases (Simpson, McCauley, Roland, Ruth, & Williges, 1985). An optimum phrase size appears to be between four and eight syllables. This message length is assumed to optimize listener attention and message intelligibility. We have taken these findings as the guidelines for the study and used the standard syntax of a general class of objects, followed by the specific item of concern, followed by the problem. In our studies, participants had to listen to the warning message generated in a fictitious car. The selected warning message was three words: "break oil leakage."

All the parameters investigated in Experiment 1 had two levels and the specifics of each of these levels can be seen in Table 1. Figure 1 shows the two types of fundamental frequency (F_0) contour and the waveform of the stimulus. These F_0 contours were obtained by a SIFT (simplified inverse filter tracking) algorithm based upon linear prediction principles (Markel & Gray, 1976). The presentation order of the 32 stimuli (16 conditions, two replication) was individually randomized for each participant by the computer.

TABLE 1. Two Levels of Five Parameters Used in Experiment 1

Parameter	Level
Participant Group	Native English speakers Non-native English speakers
Voice Type	Male-sounding Female-sounding
Speech Rate	140 words/min 200 words/min
Average F_0	125 Hz 255 Hz
F_0 Contour Type	A B

Notes. F_0—fundamental frequency.

3.1.4. Procedure

Participants were tested individually in a quiet room. Each participant was told the general nature of the study and asked to read the

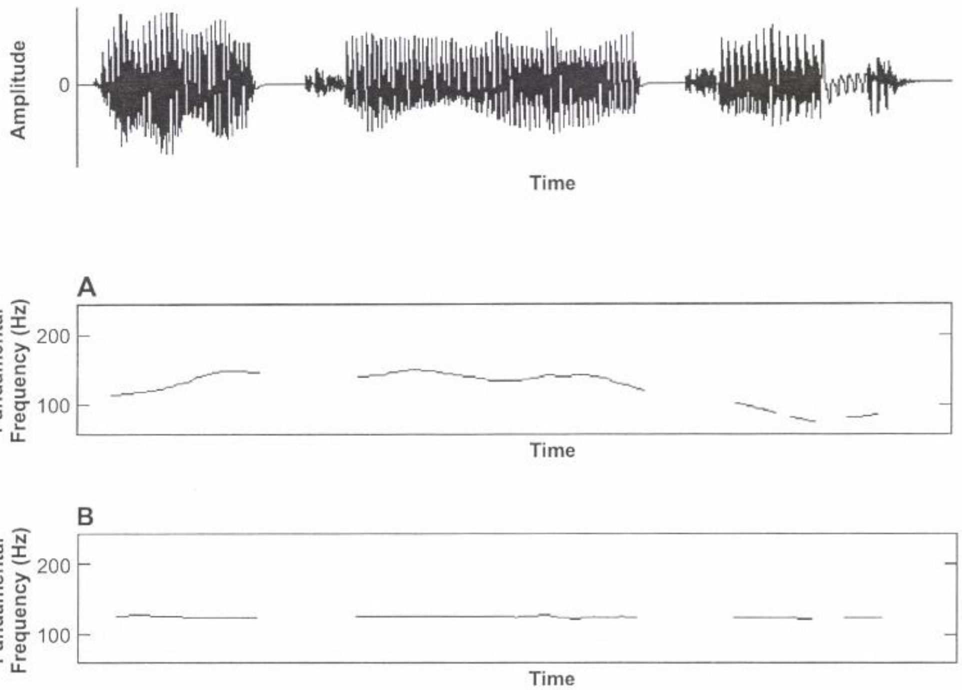

Figure 1. The waveform and two types of F_0 contour used in Experiment 1.

typewritten instructions. The participants were trained to recognize the synthesized voice warning messages and were familiarized with the experimental procedure. When the participant was ready, the warning message was played and the participant was asked to estimate the urgency of the stimulus by drawing a line and assigning a number, so that its subjective length and value were equal to the subjective urgency of the stimulus. The next stimulus was played when the participant had completed the judgement and clicked the "next" button on the computer screen. This sequence was repeated until the 16 stimuli had each been heard twice in random order. When all the participants had finished, their comments on the study were recorded.

3.2. Results and Discussion

The subjective perceived urgency of a stimulus was scaled with two response modalities: line production (*LP*) and numeric estimation (*NE*). Lodge (1981) proposed a correction formula, which is simple and easy to apply and which showed that the formula gives a better approximation of the "true" scale values by correcting for regression bias. We used this formula to get the integrated measure of the subjective perceived urgency and correct for bias in measures. The correction formula used to obtain the integrated measure (*IM*) of perceived urgency is

$$IM = (NE^{1/m_N} \times LP^{1/m_L})^{0.5},$$

where m_N is the exponent of each participant uncovered from the *NE* matches made to the line length stimuli in the pretest, and m_L is the exponent relating *LP* to the number stimuli.

A 2 (Participant Group: native vs. non-native English speakers) × 2 (Average F_0) × 2 (Speech Rate) × 2 (Voice Type) × 2 (F_0 Contour) mixed model analysis of variance (ANOVA) with repeated measures on four factors was performed. Participant Group was a between-participants variable and the integrated measure (*IM*) was used for the dependant variable in the ANOVA.

The result showed significant main effects for Average F_0, $F(1,18) = 33.59$, $p < .0001$; Speech Rate, $F(1,18) = 164.87$, $p < .0001$; Voice Type, $F(1,18) = 26.08$, $p < .0001$; and F_0 Contour, $F(1,18) = 22.28$, $p < .0002$. There were two significant interactions. First, a significant Average F_0 × Participant Group interaction, $F(1,18) = 7.42$, $p < .0139$, was found. The F_0 Contour × Participant Group interaction was also significant, $F(1,18) = 6.19$, $p < .0229$. No other significant interactions between parameters were obtained. Tables 2 and 3 show the mean urgency estimations of each level of the parameters tested. The perceived urgency of Tables 2 and 3 is the relative value when the perceived urgency for a stimulus that has a speech rate of 140 words/min, a female-sounding voice, a F_0 contour type A, and an average F_0 of 125 Hz is 1.0.

The result showed that perceived urgency is higher in a male-sounding voice than in a female-sounding one. One explanation for this may be that differences in smoothness and timbre are supposed to have affected the perceived urgency. The smoothness and timbre of the synthesized voice remained uncontrolled in this study. Some of the participants commented that the male-sounding voice was harsher than the female-sounding one.

The results also showed there were differences in rating the perceived urgency of parameters related F_0 (average F_0, F_0 contour type) between the two participant groups. One explanation for this may be that the characteristics of the mother tongue of each group are different. The warning messages used in this study were synthesized in American English. In the English language, stress plays an important role in representing the meaning of a word, but in the Korean language, stress is used only for emphasizing a specific word (Bae, 1996). Hakkinen and Williges (1984) found that if voice warnings are used, but there are many other uses of voice messages in that same environment, then the response to such warnings decreases dramatically. In order to alert the listener, voice warnings would need to stand out from the background speech (Barber et al., 1992). One easy method to make a synthesized voice distinctive from human speech would be to vary its parameters. However, according to our findings, this would affect the perceived urgency of the situations being signaled and the listener could come to regard the inappropriate urgent warnings as "crying wolf." Further, the results indicate that it is possible to achieve urgency mapping on synthesized voice warnings.

TABLE 2. Mean Urgency Estimations (and *SD*) as a Function of Voice Type and Speech Rate

Parameter	Level	All Participants (*SD*)
Voice Type	Male-sounding	1.69 (0.74)
	Female-sounding	1.39 (0.67)
Speech Rate	140 words/min	1.14 (0.54)
	200 words/min	1.94 (0.65)

TABLE 3. Mean Urgency Estimations (and *SD*) as a Function of Average F_0 and F_0 Contour Type

		Participant Group	
Parameter	Level	Native English Speakers (*SD*)	Non-Native English Speakers (*SD*)
Average F_0	125 Hz	1.19 (0.51)	1.50 (0.60)
	255 Hz	1.40 (0.64)	2.07 (0.78)
F_0 Contour Type	A	1.26 (0.57)	1.67 (0.77)
	B	1.33 (0.60)	1.90 (0.73)

Notes. F_0—fundamental frequency.

4. EXPERIMENTS 2–3

In Experiment 1, the variations in average fundamental frequency (F_0), speech rate, voice type, and fundamental frequency contour caused significant variations in the perceived urgency of the warning message. However, the results give only information about the basic relations between the variations. The main objective of Experiments 2 and 3 was to determine the power of the relations between objective changes in each of the quantifiable parameters such as average F_0 and speech rate, and the observed changes in the perceived urgency. The results of these experiments are expressed in terms of Stevens's equation that relates objective and subjective measures. Experiment 1 has showed that there were no significant interactions between the parameters of synthesized voice. Therefore, in Experiments 2 and 3, a single quantifiable parameter was varied, whereas all the others were held constant, and the participants judged the subjective urgency of the stimuli by magnitude estimation.

4.1. Method

4.1.1. Participants and apparatus

The participants and apparatus were identical to those in Experiment 1.

4.1.2. Stimuli

The warning message was as described in Experiment 1. In Experiment 2, stimuli varied in average F_0 and this parameter had six levels. All stimuli had a speech rate of 160 words/min, a male-sounding voice, and a F_0 contour type A (see Figure 1). In Experiment 3, stimuli varied in speech rate and this variable also had six levels. All six stimuli had an average F_0 of 125 Hz, a male-sounding voice, and a F_0 contour type A. The values of F_0 and speech rate varied in Experiments 2 and 3 are listed in Table 4.

In both experiments, the presentation order of the 12 stimuli (six levels, two replication) was fully randomized for each participant.

TABLE 4. Six Levels of Two Variables Used in Experiments 2 and 3

Level	Average F_0 (Hz) (Experiment 2)	Speech Rate (words/min) (Experiment 3)
1	64	120
2	89	140
3	113	160
4	138	180
5	162	200
6	186	220

Notes. F_0—fundamental frequency.

4.1.3. Procedure

The procedure was identical to that of Experiment 1 except that six stimuli were used in each experiment.

4.2. Results and Discussion

A 2 (Participant Group) × 6 (Average F_0) mixed model analysis of variance (ANOVA) with repeated measures on one factor was conducted in Experiment 2. The same correction method as in Experiment 1 was used to obtain the integrated measures. The result showed significant main effects for Average F_0, $F(5,14) = 12.10$, $p < .0001$; Average F_0 × Participant Group interaction, $F(5,14) = 3.04$, $p < .0463$. In Experiment 3, a 2 (Participant Group) × 6 (Speech Rate) mixed model analysis of variance (ANOVA) with repeated measures on one factor was conducted. The main effect of Speech Rate was significant, $F(5,14) = 17.41$, $p < .0001$, and no other significant interactions were obtained. These results were identical to those obtained in Experiment 1. The contrast analysis of both Experiments 2 and 3 showed significant differences (in all cases $p < .0463$) for all pairwise comparisons between the six levels.

The geometric means of the responses were calculated for each level and regression analysis was used to fit the data to a straight line in log-log coordinates. Figures 2 and 3 show the relations between each of the two parameters and the perceived urgency of the stimuli. The perceived urgency in Figure 2 is the relative value when the perceived urgency for a stimulus that has a speech rate of 160 words/min, a male-sounding voice, a F_0 contour type A, and an average F_0 of 113 Hz is 1.0. The perceived urgency in Figure 3 is for a stimulus of the same conditions, except for an average F_0 of 125 Hz. The results show that increases in average F_0 and speech rate result in increases in the perceived urgency of the spoken messages.

The exponent (m) of the function relating fundamental frequency (F_0) and perceived urgency was 0.879 and the coefficient of determination was .989 for native English speakers. The exponent (m) was 1.057 and the coefficient of determination was .971 for non-native English speakers. Exponents larger than 1 imply that it takes relatively small changes to produce a unit change in perceived urgency. The larger the exponent, the greater the change in perceived urgency can be produced by a change in the parameter. That is, non-native English speakers are considered to be more sensitive to the changes in fundamental frequency than native English speakers. For speech rate, the exponent (m) was 1.687 and the coefficient of determination was .998 for all participants. In both experiments, Stevens's (1966) prediction that the matching function would be a straight line in a log-log plot was supported. Besides, the cross-modality matching paradigm was used to validate the application of Stevens's power law (1957). When the *LP* response was plotted against the *NE* response to each stimulus in log-log coordinates, the function was linear, with a correlation of .980, and a slope of 1.018. This value falls within 95% confidence limits constructed around the theoretical slope of 1.0. Thus, the changes in perceived urgency with increases in parameter level are systematic and quantifiable.

The exponent can also be used to predict with the use of Stevens's power law (1957) the relative amount of change in the perceived urgency when the value of the parameter changes. For example, it is possible to calculate the amount of change in a parameter to produce a 50% increase, a doubling, and a tripling in the perceived urgency. In practical terms, it is also important to be able to equate changes in urgency across different parameters as Hellier et al. (1993) pointed out. Such information could be used to create sets of warnings that communicate the same level of urgency through different parameters. Table 5 shows the values by which each parameter must be increased to effect the specified increments in perceived urgency. For example, to double the perceived urgency of a certain warning for native English speakers,

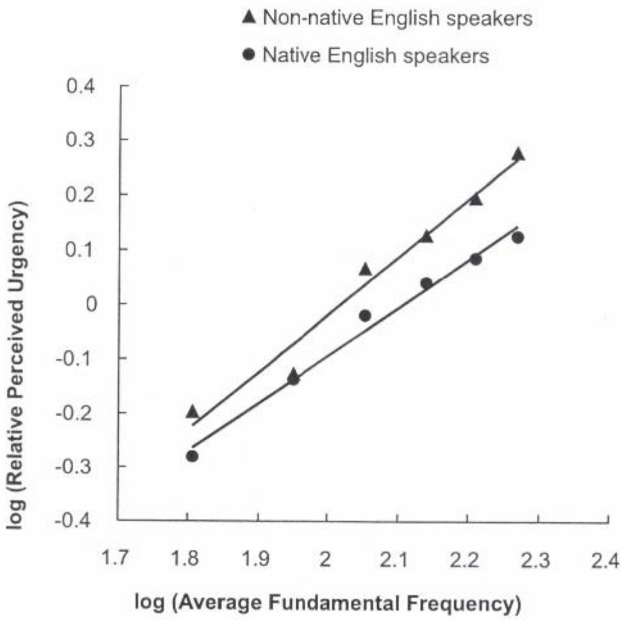

Figure 2. Relationship between average F_0 and perceived urgency.

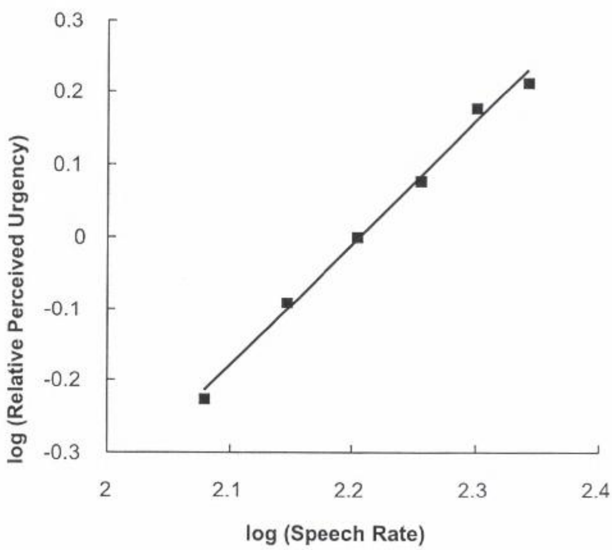

Figure 3. Relationship between speech rate and perceived urgency for all subjects.

either the average F_0 would have to be increased 220% or the speech rate would have to be increased 150.8%. Each of these changes should theoretically produce the same change in perceived urgency. The property of Stevens's power law allows calculation for only one parameter, hence these individual predictions were made for each parameter.

TABLE 5. Relation Between F_0 and Speech Rate and Perceived Urgency

Parameter	Participant Group	Exponent	Increment to Increase Perceived Urgency		
			50%	200%	300%
Average F_0	Native English speakers	0.879	45.4%	220.0%	348.9%
	Non-native speakers	1.057	51.9%	192.6%	282.7%
Speech Rate	All participants	1.687	66.3%	150.8%	191.8%

Notes. F_0—fundamental frequency.

5. EXPERIMENT 4

In Experiment 1, the variations in fundamental frequency (F_0) contour caused significant variations in perceived urgency. Studies of the effects of emotion on the acoustic characteristics of human speech have also shown that the F_0 contour is one of the most sensitive indicators of the emotional state of the speaker (Williams & Stevens, 1972). Hence, in Experiment 4, F_0 contour was selected for detailed analysis. Five types of F_0 contour were constructed with other parameters kept constant. In addition, a comparison between the urgency ratings of participant groups—native English speakers versus non-native English speakers (Koreans)—was made. Participants used the magnitude estimation task to judge the perceived urgency of the stimuli as described in previous experiments.

5.1. Method

5.1.1. Participants and apparatus

The participants and apparatus were identical to those of Experiment 1.

5.1.2. Stimuli

The warning message used in Experiment 4 was the same as in previous experiments. Five types of F_0 contour were used whereas all other

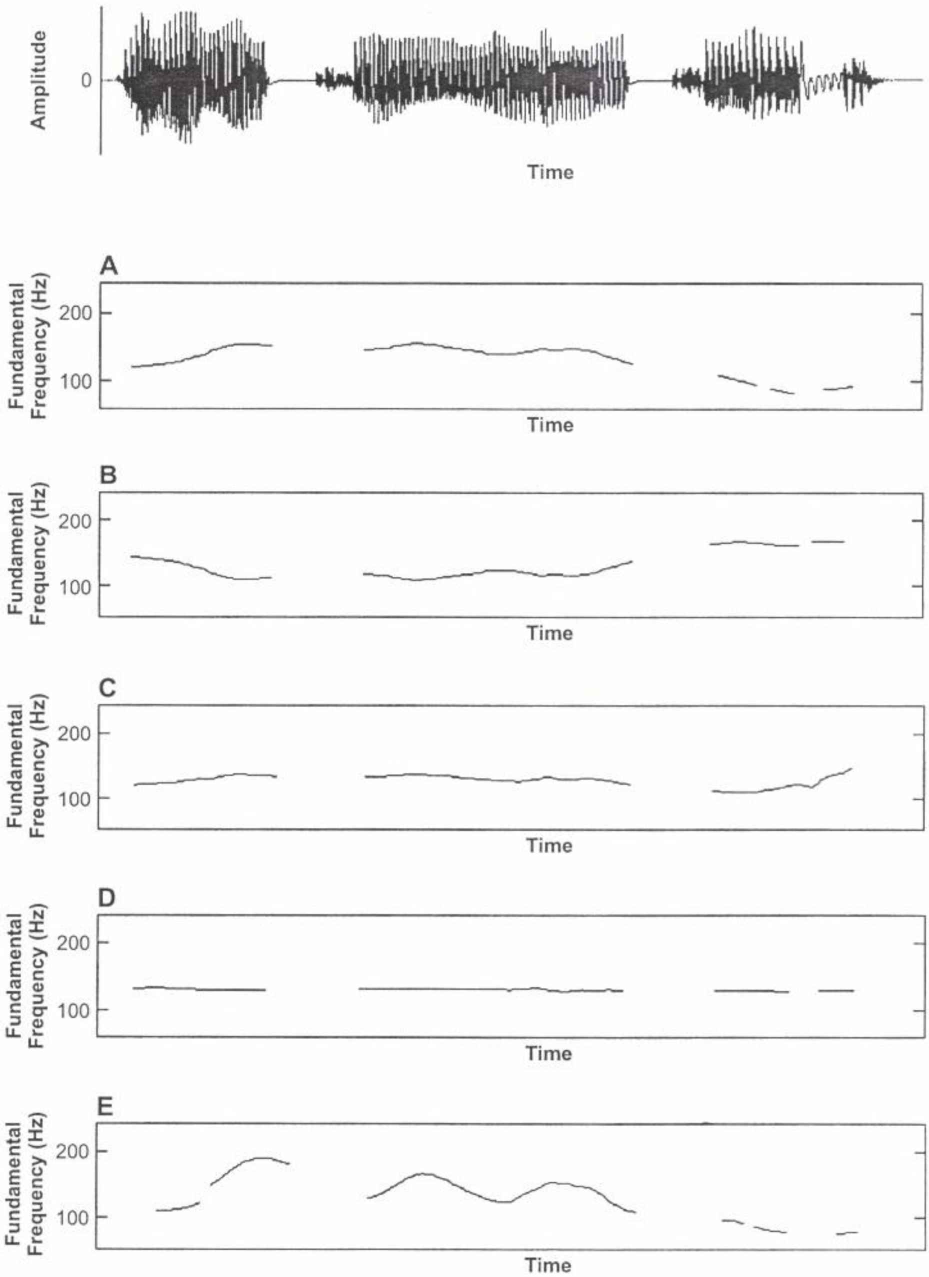

Figure 4. The waveform and five types of F_0 contour used in Experiment 4.

parameters of the voice were held constant. The stimuli had a speech rate of 175 words/min, a male-sounding voice, and an average F_0 of 131 Hz. Figure 4 shows the five types of F_0 contour and the waveform of the stimulus. F_0 contour A sounds like normal speech or reading a text and no modification was applied to this contour. The contour has slow fluctuations and a slow falling in the final part of the phase. F_0 contour B was created by reversing F_0 contour A. F_0 contour C has an interrogatory inflection. The contour shape is relatively flat with few fluctuations and F_0 is rising toward the end. The shape of F_0 contour D is flat and monotonous. F_0 contour E was created by doubling the fluctuations of F_0 contour A and thus it has rapid up-and-down fluctuations within the voiced interval. F_0 contours A and D were the same as those used in Experiment 1. The presentation order of the 10 stimuli (five levels, two replication) was fully randomized for each participant.

5.1.3. Procedure

The procedure was identical to that of Experiment 1 except that five stimuli were used in this experiment.

5.2. Results and Discussion

The same correction method as the one in Experiment 1 was used to obtain the integrated measures. A 2 (Participant Group: native vs. non-native English speakers) × 5 (F_0 Contour types) ANOVA with repeated measures on one factor was performed using perceived urgency as the dependent variable. The result showed significant main effects for both Participant Group, $F(1,18) = 6.19$, $p < .022$, and F_0 Contour type, $F(4,15) = 8.11$, $p < .0011$. Figure 5 shows the mean urgency ratings for each F_0 contour type and participant group. The perceived urgency in Figure 5 is the relative value when the perceived urgency for a stimulus that has a speech rate of 175 words/min, a male-sounding voice, an F_0 contour A, and an average F_0 of 131 Hz is 1.0. In addition, contrast analysis was conducted to obtain more detailed information about the differences among the mean urgency ratings of F_0 contour types in each participant group. The results showed that the native English speakers rated F_0 contour B significantly less urgent ($p < .001$) than other types of F_0 contour. The other F_0 contour types did not significantly differ in the ratings by the native English speakers. The results of the non-native English speakers (Koreans) showed significant differences for all pairwise comparisons ($p < .05$) except between F_0 contours A and C. The non-native English speakers rated F_0 contour D the most urgent, and B the least urgent. The comparisons of urgency ratings between the two participant groups showed that the native English speakers rated F_0 contours B and D significantly less urgent ($p < .05$) than the non-native English speakers did. The non-native English speakers rated the other F_0 contour types slightly but not significantly more urgent than the native English speakers did.

The results indicated that, for native English speakers, the various types of F_0 contour could be used to create new warning sounds so as to make them distinctive from background speech, while retaining the meaning and the level of urgency. However, this application may not be practicable for non-native English speakers except to use F_0 contour types A and C. These results corresponded with the assertion that the stress and intonation are used for emphasizing specific words in the Korean language (Bae, 1996). Further, this implies that it is possible to generate various levels of perceived urgency by altering the F_0 contours for non-native English speakers (Korean).

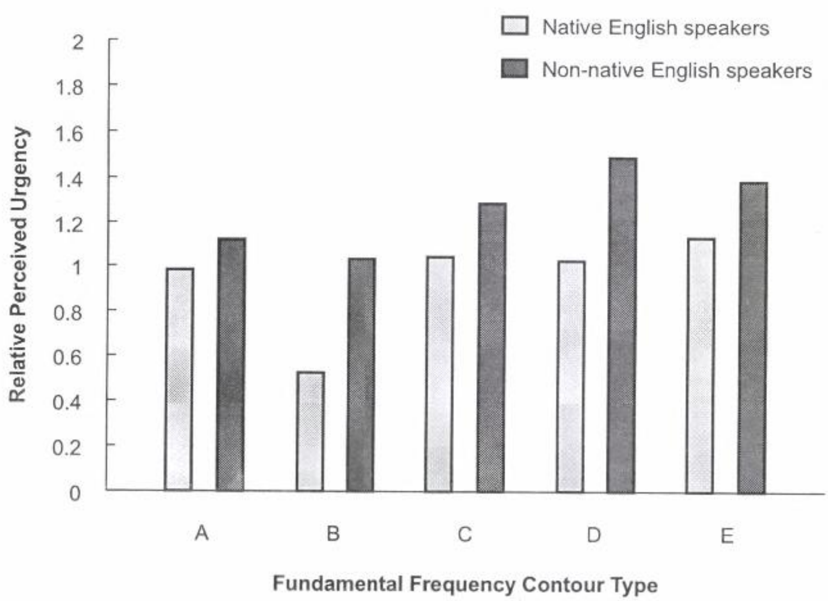

Figure 5. Mean perceived urgency for native and non-native English speakers.

6. GENERAL DISCUSSION

The purpose of this study was to provide basic data regarding the effects of various operational parameters of synthesized speech on the perceived urgency of warning signals. The results showed that all investigated acoustic parameters, speech rate, average fundamental frequency (F_0), voice type, and fundamental frequency contour have clear effects on the subjective perceived urgency of synthesized voice warnings. For two quantitative parameters, speech rate and average F_0, the results showed that it is possible to use psychophysical techniques (Stevens's power law, 1957) to quantify the effects of acoustic changes on perceived urgency. The speech rate had a higher exponent than average F_0 in Stevens's power function, that is, it took smaller changes in speech rate than in average F_0 to produce specified changes in perceived urgency.

These results could be used to generate various levels of perceived urgency by altering the parameters of synthesized voice warnings. Further, they could be applied in designing advanced synthesized voice warning systems and in improving existing warning systems. It should be possible to match synthesized voice warning sounds to the subjective urgency of the situation if the results and the predictions of the quantitative parameters are properly used.

Barber et al. (1992) argued that synthesized speech could be a very useful means of "alarm display" by being easily distinguishable from human speech. If a synthesized voice sounds too human, it may cause problems. It may lead the listener to attribute too much intelligibility to the device and may be confused with background human speech (Cotton, McCauley, North, & Streib, 1983). The U.S. Department of Defense (1981) also proposed that synthesized speech should sound distinctive from human speech. The results of this study showed that the changes in urgency could be equated across different parameters such as speech rate and average F_0. This information could be used to create sets of synthesized voice warning sounds that communicate the same level of urgency with sounds with different parameters. Further research on the intelligibility and the effectiveness of synthesized speech messages is needed. As this study was conducted in quiet and was limited to a single warning phrase, the results may not hold in noise and for other type of warning phrases. However, the results of this study could form the basis for such research and provide guidelines for new warning systems that sound distinguishable from human speech while retaining the high quality of synthesized voice.

REFERENCES

Aretz, A.J. (1983). A comparison of manual and vocal response models to the control of aircraft systems. In *Proceedings of the 27th Annual Meeting of the Human Factors Society* (pp. 97–101). Santa Monica, CA: Human Factors Society.

Bae, J.C. (1996). *Introduction to Korean phonology*. Seoul, Korea: Shingu.

Barber, C., Stanton, N.A., & Stockley, A. (1992). Can speech be used for alarm displays in "process control" type tasks? *Behavior & Information Technology, 11*, 216–226.

Byblow, W.D., & Corlett, J.T. (1989). Effects of linguistic redundancy and coded voice warnings on system response time. *Applied Ergonomics, 20*, 105–108.

Cotton, J.C., McCauley, M.E., North, R.A., & Streib, M.I. (1983). *Development of speech input/output for tactical aircraft* (Report No. AFWAL-TR-83-3073). Dayton, OH: U.S. Air Force.

Edworthy, J. (1994). The design and implementation of non-verbal auditory warnings. *Applied Ergonomics, 25*(4), 202–210.

Edworthy, J., Loxley, S., & Dennis, I. (1991). Improving auditory warning design: Relationship between warning sound parameters and perceived urgency. *Human Factors, 33*, 205–231.

Gescheider, G.A. (1985). *Psychophysics—Method, theory, and application*. London: Laurence Erlbaum Associates.

Hakkinen, M.T., & Williges, B.H. (1984). Synthesized warning messages: Effects of an alerting cue in single-and multiple-function voice synthesis systems. *Human Factors, 26*(2), 185–195.

Hellier, E.J., Edworthy, J., & Dennis, I. (1993). Improving auditory warning design: Quantifying and predicting the effects of different warning parameters on perceived urgency. *Human Factors, 35*(4), 693–706.

Irving, A. (1981). *An experimental comparison of operator responses to voice and tone system warnings* (Report No. BAE-BT-12051). Bristol, UK: British Aerospace Dynamic Group.

James, S.H., & James, M.R. (1989). The effects of warning format and keyboard layout on reaction times to auditory warnings. *Proceedings of the Institute of Acoustics, 11*(5), 25–42.

Larkey, L.S., & Danly, M. (1983, November). *Fundamental frequency and the comprehension of simple and complex sentences*. Paper presented at the 106th meeting of the Acoustical Society of America, San Diego, CA.

Lodge, M. (1981). *Magnitude scaling: Quantitative measurement of opinions*. Beverly Hills, CA: Sage.

Markel, J.D., & Gray, A.H., Jr. (1976). *Linear prediction of speech*. New York: Springer-Verlag.

Otsubo, S.M. (1988). A behavioral study of warning labels for consumer products: Perceived danger and use of pictographs. In *Proceedings of the 32nd Meeting of the Human Factors Society* (pp. 536–540). Santa Monica, CA: Human Factors Society.

Simpson, C.A., & Marchionda-Frost, K. (1984). Synthesized speech rate and pitch effects on intelligibility of warning messages for pilots. *Human Factors, 26*, 509–517.

Simpson, C.A., McCauley, M.E., Roland, E.F., Ruth, J.C., & Williges, B.H. (1985). System design for speech recognition and generation. *Human Factors, 27*(2), 115–141.

Simpson, C.A., McCauley, M.E., Roland, E.F., Ruth, J.C., & Williges, B.H. (1987). Speech controls and displays. In G. Salvendy (Ed.), *Handbook on human factors* (pp. 1490–1525). New York: Wiley.

Simpson, C.A., & Williams, D.H. (1980). Response time effects of alerting tone and semantic context for synthesized voice cockpit warnings. *Human Factors, 22*(3), 319–330.

Stevens, S.S. (1955). The measure of loudness. *Journal of the Acoustical Society of America, 27*, 815–820.

Stevens, S.S. (1957). On the psychophysical law. *Psychological Review, 64*, 153–181.

Stevens, S.S. (1958). Problems and methods of psychophysics. *Psychological Bulletin, 55*, 177–196.

Stevens, S.S. (1966). Matching function between loudness and ten other continua. *Perception & Psychophysics, 1*, 5–8.

Stevens, S.S., Carton, A.S., & Shickman, G.M. (1958). A scale of apparent intensity of electric shock. *Journal of Experimental Psychology, 56*, 328–334.

Stevens, S.S., & Guirao, M. (1963). Subjective scaling of length and area and the matching of length to hardness and brightness. *Journal of Experimental Psychology, 59*, 60–67.

Stevens, J.C., & Stevens, S.S. (1963). Brightness function: Effects of adaptation. *Journal of the Optical Society of America, 53*, 375–385.

Streeter, L.A., Macdonald, N.H., Apple, W., Krauss, R.M., & Galotti, K.M. (1983). Acoustic and perceptual indicators of emotional stress. *Journal of the Acoustical Society of America, 73*, 1354–1360.

U.S. Department of Defense. (1981). *Human engineering design criteria for military systems, equipment and facilities* (Report No. MIL-STD-1472C). Washington, DC: Author.

Wheale, J.L. (1983). Evaluation of an experimental central warning system with a synthesized voice component. *Aviation, Space, and Environmental Medicine, 54*, 517–523.

Williams, C.E., & Stevens, K.N. (1972). Emotions and speech: Some acoustical correlates. *Journal of the Acoustical Society of America, 52*, 1238–1249.