

PAPER

Verbal disaster warnings and perceived intelligibility, reliability, and urgency: The effects of voice gender, fundamental frequency, and speaking rate

Kenta Ofuji^{1,*} and Naomi Ogasawara^{2,†}

¹*Department of Computer Science and Engineering, University of Aizu,
90 Kami-Iawase, Tsuruga, Ikki-machi, Aizuwakamatsu, 965–8580 Japan*

²*Department of International Communication, Gunma Prefectural Women's University,
1395–1 Kaminote, Tamamura-machi, Gunma, 370–1193 Japan*

(Received 27 March 2017, Accepted for publication 12 October 2017)

Abstract: In this paper, we study the effects of acoustic characteristics of spoken disaster warnings in Japanese on listeners' perceived intelligibility, reliability, and urgency. Our findings are threefold: (a) For both speaking speed and f_0 , setting them to normal (compared from slow/fast (+/–20%) for speed, and from low/high (+/– up to 36 Hz) for f_0) improved the average evaluations for Intelligibility and Reliability. (b) For Urgency only, setting speed to faster (both slow to normal and normal to fast) or setting f_0 to higher (both low to normal and normal to high) resulted in an improved average evaluation. (c) For all of intelligibility, reliability, and urgency, the main effect of speaking speed was the most dominant. In particular, urgency can be influenced by the speed factor alone by up to 39%. By setting speed to fast (+20%), all other things being equal, the average perceived urgency raised to 4.0 on the 1–5 scale from 3.2 when the speed is normal. Based on these results, we argue that the speech rate may effectively be varied depending on the purpose of an evacuation call, whether it prioritizes urgency, or intelligibility and reliability. Care should be taken to the possibility that the respondent-specific variation and experimental conditions may interplay these results.

Keywords: Verbal warning, Acoustic parameters, Perception

PACS number: 43.71.Gv [doi:10.1250/ast.39.56]

1. INTRODUCTION

1.1. Responses to Disaster Warnings

When a natural disaster occurs, people go through a process before beginning evacuation behaviors: (1) recognizing the situation and disaster information; (2) making a decision about evacuation based on the provided information and their own knowledge and past experiences; and (3) taking action [1]. Recognition and decision are mental activities, and several models attempt to depict such psychological processes people go through after they receive disaster information and before they actually evacuate. Nakamura's psychological model [2] considers the recognition of danger and related social elements (e.g., urge for evacuation brought by family and neighbors) and the facilitating and prohibiting factors associated with them. Among the major factors Sorensen and Sorensen [3]

list, 'perceived risk' is a facilitating factor, which increases the likelihood of evacuation.

Nevertheless, people usually tend to deny or underestimate the possibility and damages brought by a disaster when they receive disaster information. According to Drabek [4, p. 515], "The first principle in understanding disaster warning responses is to recognize explicitly that *the initial response to any warning is denial*." Such reaction to warnings is due to a normalcy bias, a type of cognitive bias, by which one's perception or expectation is skewed from a rational judgment and often associates with an optimistic view and an outsider's perspective. When people receive a disaster warning, they tend to think "Nothing serious will happen," or "I will survive even if others die," without any logical reasoning [5–7].

A low evacuation rate in many of the devastated areas in Japan has been reported, and people's low evacuation awareness has been considered a serious matter [8]. In an attempt to increase the evacuation rate, some approaches for making evacuation calls more effective have been

*e-mail: o-fu@u-aizu.ac.jp

†e-mail: naomi-o@fic.gpwu.ac.jp

discussed among researchers, disaster management practitioners, and relevant ministries and agencies. The elements in verbal warnings, such as wording, tone of command, context, and acoustics, are all subjects under re-examination [9,10].

There have been many studies on ‘what to say’ in verbal warnings, but less amount of research has paid attention to ‘how to say’ warnings. This study sheds light on the effects of acoustic characteristics of spoken warnings on listeners’ perceived intelligibility, reliability, and urgency. We hope that our findings will be beneficial to researchers and practitioners in the field of disaster management.

1.2. Acoustic Correlates with Perception of Warning

It has been examined and shown that suprasegmentals are closely related with how speech conveys human emotions. Murray and Arnott [11] review the previous literature on voice and emotions. They summarize how vocal features such as f_0 , duration, loudness, and timbre, in speech are affected by emotions, and listeners can recognize speaker’s emotions through speech within a very short period of time.

On the perception side, acoustic parameters like timbre, speed, loudness, f_0 , and harmonics arouse emotions of the listener as well. A number of auditory experiments have been conducted to examine the level of acoustic parameters of sound alerts for enhancing perceived urgency [12]. Since the principal roles of a warning are attracting listener’s attention and compelling immediate reactions, arousing listener’s feeling of urge through perceived urgency is critical. In the case of auditory signals such as pure tones, beeps, and bells, it was found that a signal with a higher frequency (500 Hz and 800 Hz), a higher level of loudness (40 dB), and no inter-pulse interval was perceived as the most urgent and obtained the shortest reaction time (RT) [13]. Suied, Susini, and McAdams also found that a shorter interonset interval (100 ms) of 1,000 Hz pure tone obtained a higher urgency rating and faster RT compared with 300 ms interonset interval of the same signal [14]. As well as non-verbal sound alerts, acoustic characteristics of verbal warnings were found to be effective on perception, and the manipulation of f_0 , speed, and loudness can alter perceived urgency in a similar fashion to non-verbal warnings [15]. Park and Jang [16] used synthesized voice warnings for their investigation and found that a higher perceived urgency was obtained with a warning that was male-sounding, faster (200 words/min), had a higher average f_0 (255 Hz), and also had a more dynamic f_0 contour. Regarding the voice gender, other experiments revealed the opposite result that female voice with higher f_0 is more effective on perceived urgency and preferred by listeners than a male voice [15,17,18].

The findings regarding the superiority of a female voice is attributed to a higher f_0 than that of male voice, from which it can be concluded that f_0 is definitely an essential parameter for perceived urgency. In addition to f_0 , we think that speaking rate is also an important parameter for boosting urgency. Moreover, we believe that urgency is not the only driving force to increase the efficiency of verbal warnings. Warnings have to be understood easily in order to quickly deliver a message to the listeners, and they have to be trustworthy in order to propel evacuation behaviors. In this sense, perceived intelligibility and reliability are also important elements for designing verbal warnings.

2. PERCEPTION EXPERIMENT

2.1. Stimuli

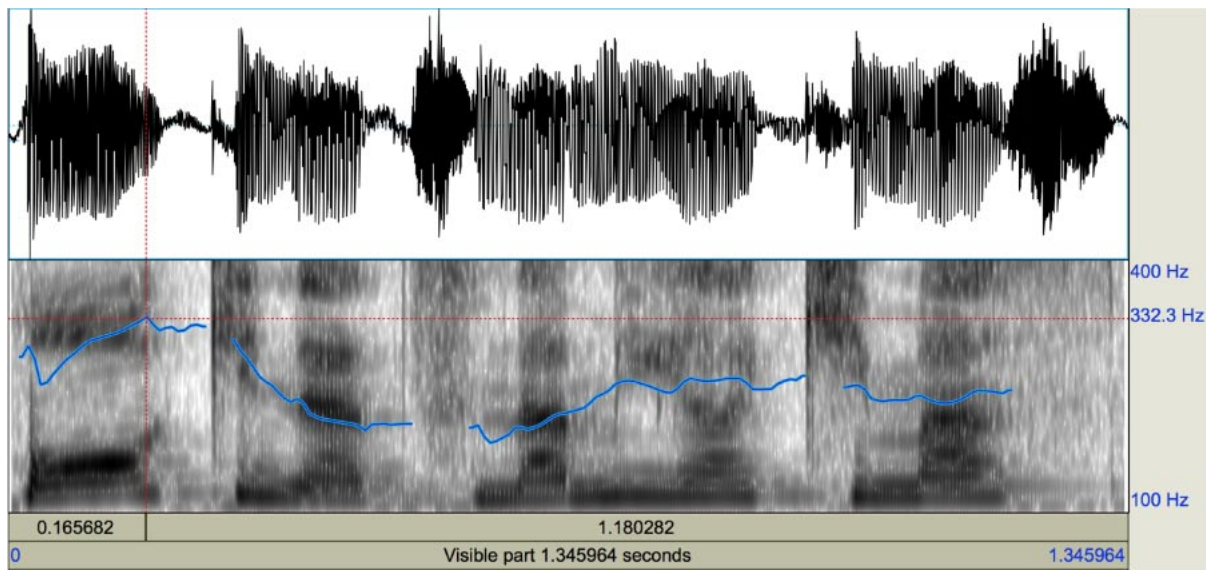
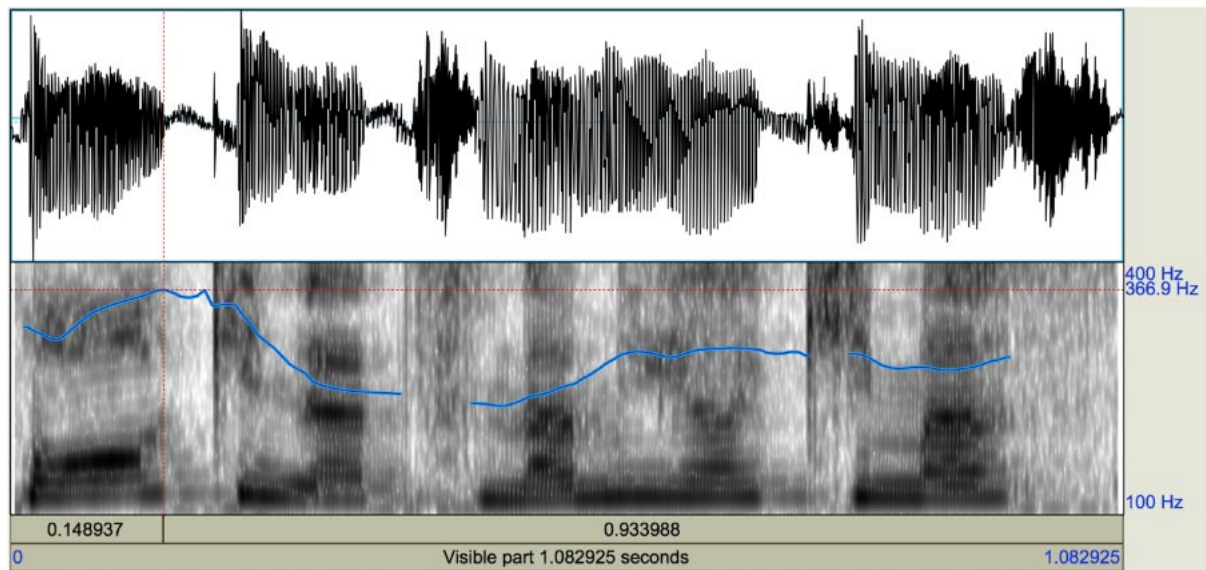
Three short warnings in Japanese were created: 1. *Kyoo wa ame ga furimasu. Kasa wo motte dekakete kudasai.* ‘It’s going to rain today. Please go out taking an umbrella with you.’ 2. *Ookina tsunami ga kimasu. Tadachini hinan shitekudasai.* ‘A big tsunami is coming. Please evacuate immediately.’ and 3. *Gakekuzure no kiken ga arimasu. Tadachini hinan shitekudasai.* ‘There is a risk of landslide. Please evacuate immediately.’ Warning 1 was used during the practice phase prior to the test phase in order the participants to familiarize with the task. Warnings 2 and 3 were used for the test phase. Each warning is composed of two short sentences with 24 morae (Warnings 1 and 2) or 27 morae (Warning 3). A female (in her 40’s) and a male (in his late 20’s) native speaker of Japanese, who both have relatively clear voices and articulations, read the warnings out aloud with their normal speed in a sound attenuated booth, and their utterances were recorded with a high quality condenser microphone (sE2200aII) and recording device (Avid, MBox Pro with ProTools 9), sampled at 44.1 kHz and 16 bit.

Because the volume of the male utterances was not as much as the female utterances, the average sound pressure level of each male utterance was raised using PRAAT software [19] in order to match with the female utterances. The average SPL of each of the utterances was adjusted to 73–74 dB on the software. The SPL-matched tokens were named ‘normal,’ indicating no further modification was made for them. See Table 1 for the acoustic information about the normal stimuli.

Next, speaking speed and f_0 were modified with the normal tokens. Tokens that were 20% faster and slower and 18–36 Hz raised and lowered in average than the normal tokens were created using PRAAT [19]. Figure 1 illustrates a female normal token and the same token with raised f_0 and faster speed as samples. The modification level was set where two native Japanese speakers (including one of the authors) were able to detect the acoustic change from the

Table 1 Acoustic data of normal stimuli.

Warning type	Speaker	Duration (s)	Speed (s)	Mean f_0 (Hz)	Mean intensity (dB)
Warning 1 24 moras	Male	4.8	0.2/mora	157.3	73.8
	Female	5.6	0.23/mora	226.5	74.1
Warning 2 24 moras	Male	5.3	0.22/mora	156.2	72.8
	Female	5.1	0.21/mora	223.9	74.0
Warning 3 27 moras	Male	5.4	0.2/mora	150.9	74.1
	Female	5.4	0.2/mora	225.3	73.9

(a) Normal speed (1.34 s) — normal f_0 (highest f_0 332.3 Hz).(b) Faster speed (1.08 s) — higher f_0 (highest f_0 366.9 Hz).**Fig. 1** Token of female “Ookina Tsunami ga kimasu”.

normal ones. The total number of tokens created including the normal ones was 54 (3 warning type \times 2 gender \times 3 speaking speed \times 3 f_0). Among 18 tokens of Warning 1, only 4 tokens were used as practice stimuli, hence the number of stimuli played during the experiment became 40 (4 of Warning 1 for the practice phase and 36 of Warnings 2 and 3 for the test phase).

It may be worth noting why we chose the variation of speed to be $\pm 20\%$ and that of f_0 up to 36 Hz. In preparing the experimental setup, we explored the correct degree of these variations based both on the valid perceptual differences and on the naturalness of the resultant stimuli. That is, for example, if the speed variation is too small and subtle, it would hardly result in the perceptual difference, whereas if it is too large, it would lose the naturalness as a real-world evacuation call. We carefully checked if all the created stimuli satisfied the both conditions by asking a student assistant for an objective feedback.

2.2. Participants

Eighty-seven people were recruited for the experiment. Among them, 76 participants were students of the University of Aizu, and 11 participants, who were in their 30–60s, were either employees of the university or outsiders. Their studentship status was coded in the dataset as “Subject’s age” variable, a binary variable that indicates students/non-students. None of the participants reported having hearing or speech disorders. A small gift was given to them for their participation.

2.3. Procedures

The experiment took place in two different settings. In one setting, 34 students were sitting in a large classroom (930 cm \times 1,500 cm) and heard the pseudo-randomized stimuli through two speakers (model BOSE301v, sized 24.8 cm (H) \times 36.2 cm (W) \times 26 cm (D), equipped with a 5 cm stereo targeting tweeter, a 5 cm direct/reflecting tweeter and a 20 cm woofer) placed on the right and left front corners of the room. Another group of 42 students and 11 elderly people were tested in another setting. Two or three participants at a time sat separately in a lab (510 cm \times 750 cm) and heard the stimuli through a speaker (model BOSE SoundLink Bluetooth speaker III, 13.2 cm (H) \times 25.6 cm (W) \times 4.8 cm (D)) placed in one corner. These two different settings were prepared to simulate real-life situations when the participants would hear a disaster warning either in a classroom with many fellow students or in a small room with a few people around. Before each stimulus was played, a computer-generated voice gave a number (e.g., *1-ban* ‘No. 1’) followed by 200 ms silence. After a stimulus was played, there was 10 s silence as response window. The participants were asked to listen to

and rate each stimulus on a 1-to-5-scale (1: Strongly disagree; 2: Disagree; 3: Neither; 4: Agree; 5: Strongly agree) for each of the three perception criteria of the spoken warning.

3. RESULTS

Prior to the statistical analysis, we removed outlier data from the collected dataset. For instance, the data from the respondents who gave ‘5’ to (almost) all the stimuli were excluded from analysis. Specifically, seven respondents who had a standard deviation smaller than 0.35 across all test stimuli of intelligibility, reliability, and urgency were eliminated, leaving 80 respondents as the effective respondents, subject to the following analyses.

We applied ANOVA (ANalysis Of VAriance) to the data collected in the experiments. In the analysis, we focused on the three variables of our foremost concern: Speaker gender, Speed and f_0 (hereafter, we may call these three variables “the three stimuli”). Table 2 shows the ANOVA results for each of Intelligibility, Reliability and Urgency (hereafter, we may call these “the three evaluation criteria”). The top portion of each table shows the main effects and the interactions. In addition, each table lists the control variables, namely Subject error, Sentence type, Venue and Subject’s age. These variables control for the general experimental conditions that differed among respondents. In Table 2, in addition to the usual F test results and the associated p -values for each effect, we also show the effect sizes calculated from our sample (see Appendix for calculation methods). To interpret the results, we both look at the F test results and the sample effect sizes.

3.1. Main Effects

(a) Speed

We see from Table 2 that the main effect from Speed was the largest for all of Intelligibility, Reliability and Urgency. The associated effect sizes were 0.45, 0.58 and 1.00. These are all classified as “large” under the convention of Cohen [20]. In addition, the mean squares (MS) and the F -values of the Speed main effect were both largest in all three evaluations. This means that in our experiment, respondents perceived the magnitude of the Speed effect to be the most significant among all stimuli, for all three evaluations. It has to be noted, though, that this result also relates to our choice of the combination of the parameter ranges, where the variation in Speed was chosen to be $\pm 20\%$ of the original. Had we chosen a smaller range of the variation in Speed, the perceived magnitude of its effect may have been weakened. But at least from the experimental condition we employed this time, the Speed effect (of $\pm 20\%$) was the most prominent. It is worth stressing that, from the Urgency table in Table 2, about

Table 2 ANOVA results.

Intelligibility		<i>SS</i>	<i>d.f.</i>	<i>MS</i>	<i>F</i>	Effect size
Main effects	Speaker gender	1.7	1	1.7	2.3	0.03
	Speed	433.2	2	216.6	300.1***	0.45
	f_0	76.3	2	38.2	52.9***	0.19
Interactions	Gender \times Speed	75.1	2	37.5	52.0***	0.19
	Speed \times f_0	1.9	4	0.5	0.7	0.03
	$f_0 \times$ Gender	32.5	2	16.3	22.5***	0.12
Control variables	Subject error	706.6	81	8.7	12.1***	
	Sentence type	36.2	1	36.2	50.2***	
	Venue	45.9	1	45.9	63.7***	
	Subject's age	0.3	1	0.3	0.4	
Residuals		2,111.7	2,926	0.7		
Total		3,521.5	3,023	1.2		

Reliability		<i>SS</i>	<i>d.f.</i>	<i>MS</i>	<i>F</i>	Effect size
Main effects	Speaker gender	74.5	1	74.5	112.9***	0.20
	Speed	637.7	2	318.8	483.3***	0.58
	f_0	59.9	2	30.0	45.4***	0.18
Interactions	Gender \times Speed	31.5	2	15.7	23.9***	0.13
	Speed \times f_0	3.9	4	1.0	1.5	0.05
	$f_0 \times$ Gender	27.0	2	13.5	20.5***	0.12
Control variables	Subject error	632.1	79	8.0	12.1***	
	Sentence type	1.0	1	1.0	1.4	
	Venue	0.3	1	0.3	0.5	
	Subject's age	0.6	1	0.6	0.9	
Residuals		1,884.3	2,856	0.7		
Total		3,352.8	2,951	1.1		

Urgency		<i>SS</i>	<i>d.f.</i>	<i>MS</i>	<i>F</i>	Effect size
Main effects	Speaker gender	326.9	1	326.9	454.2***	0.39
	Speed	2140.2	2	1070.1	1486.9***	1.00
	f_0	32.1	2	16.1	22.3***	0.12
Interactions	Gender \times Speed	111.0	2	55.5	77.1***	0.23
	Speed \times f_0	3.9	4	1.0	1.4	0.04
	$f_0 \times$ Gender	8.1	2	4.1	5.7***	0.06
Control variables	Subject error	761.3	83	9.2	12.7***	
	Sentence type	19.9	1	19.9	27.6***	
	Venue	2.0	1	2.0	2.8*	
	Subject's age	4.1	1	4.1	5.7**	
Residuals		2,156.2	2,996	0.7		
Total		5,565.8	3,095	1.8		

note: *SS* = sum of squares, *d.f.* = degree of freedom, *MS* = mean squares. *** $p < 1\%$, ** $p < 5\%$, * $p < 10\%$.

39% of the variance in urgency response is influenced by the speed main effect alone (*SS* of speed 2,140.2/Total *SS* 5,565.8 = 39%). This percentage is much higher than those

for Intelligibility (433.2/3,521.5 = 12%) and Reliability (637.7/3,352.8 = 19%), implying the great influence the speaking speed can have on the perceived Urgency.

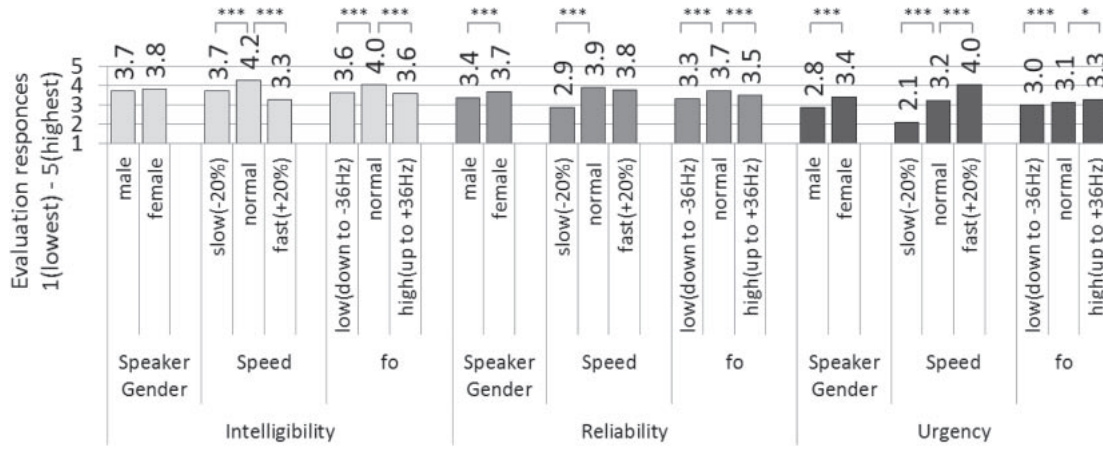


Fig. 2 Average scores under each stimulus condition and multiple comparison results. Numbers over each bar are the average evaluation scores. The asterisks and brackets over pairs of bars are statistically significant pairs of multiple comparison with the Bonferroni method. ***: $p < 0.01$, *: $p < 0.1$ and blank: $p \geq 0.1$.

(b) Speaker gender and f_0

We also see from Table 2 that, mostly, the main effect from Speaker gender and that from f_0 were behind those from Speed. For example, the effect sizes from Speaker gender were at most 0.39, not exceeding those from Speed. In particular, its effect size for Intelligibility was fairly small at 0.03, with its associated F -test result being not statistically significant. For Urgency, the effect size of 0.39 was close to the floor value of 0.40 that qualified a “large effect.” This means that setting the Speaker gender to female can have a significant effect, at least on the perceived Urgency.

As for the f_0 's main effects, all three F -test results were statistically significant, but the associated effect sizes were only in the range of 0.12 to 0.19, falling within Cohen's [20] “small effects.” We have to note, though, that this level of significance may have been due, again, to the parameter range choice of up to ± 36 Hz we took this time. Had we chosen a larger range in the frequency alteration, we may have had a different degree of response, resulting in different values in the effect sizes and the associated F -tests. Accordingly, it is worth noting that the order of magnitude where the main effect of Speed was greater than those of Speaker gender and f_0 , may have had a different look had we taken a different set of choices for the parameter alterations.

Nevertheless, under our parameter setting this time, it was confirmed that the Speed was the prominent factor that influenced all three evaluation criteria, particularly Urgency, from both the effect size and the F -test.

(c) Multiple comparisons

We performed multiple comparisons for each main effect, to discern which condition results in high evaluation score on average, for the three evaluation criteria (Fig. 2). In Fig. 2, bars show the average scores under each stimulus

condition. Numbers over each bar are the average evaluation scores. For example, the leftmost bar of 3.7 means the average evaluation of Intelligibility was 3.7 when the speaker gender was female, with all the other conditions held at the average. The asterisks and brackets over pairs of bars are statistically significant pairs of multiple comparison with the Bonferroni method (***: $p < 0.01$, *: $p < 0.1$ and blank: $p \geq 0.1$).

Figure 2 gives the following two interpretations:

- (1) When the speaker gender is female, the average evaluations for Reliability and Urgency have improved than when it is male.
- (2) For both speaking speed and f_0 , setting them to Normal (compared from slow/fast for speed and from low/high for f_0) improved the average evaluations for Intelligibility and Reliability. For Urgency only, setting speed to faster (both slow to normal and normal to fast) or setting f_0 to higher (both low to normal and normal to high) resulted in an improved average evaluation.

3.2. Interaction Effects

Among the three interactions of Speaker gender \times Speed, Speed \times f_0 , $f_0 \times$ Speaker gender, we do not step into a deep interpretation of that from Speed \times f_0 because its effect size in each evaluation criteria (Intelligibility, Reliability and Urgency) was below 0.1 or the lower limit for qualifying a “small effect” by the Cohen's convention. The other two interactions, namely Speaker gender \times Speed and $f_0 \times$ Speaker gender had, in most cases, the effect sizes that exceeded 0.1 and qualified as “small effects.” As a result, they became statistically significant in the associated F -tests.

To visually check this fact, we drew interaction plots (Fig. 3). From left panels to right, the three panels in the

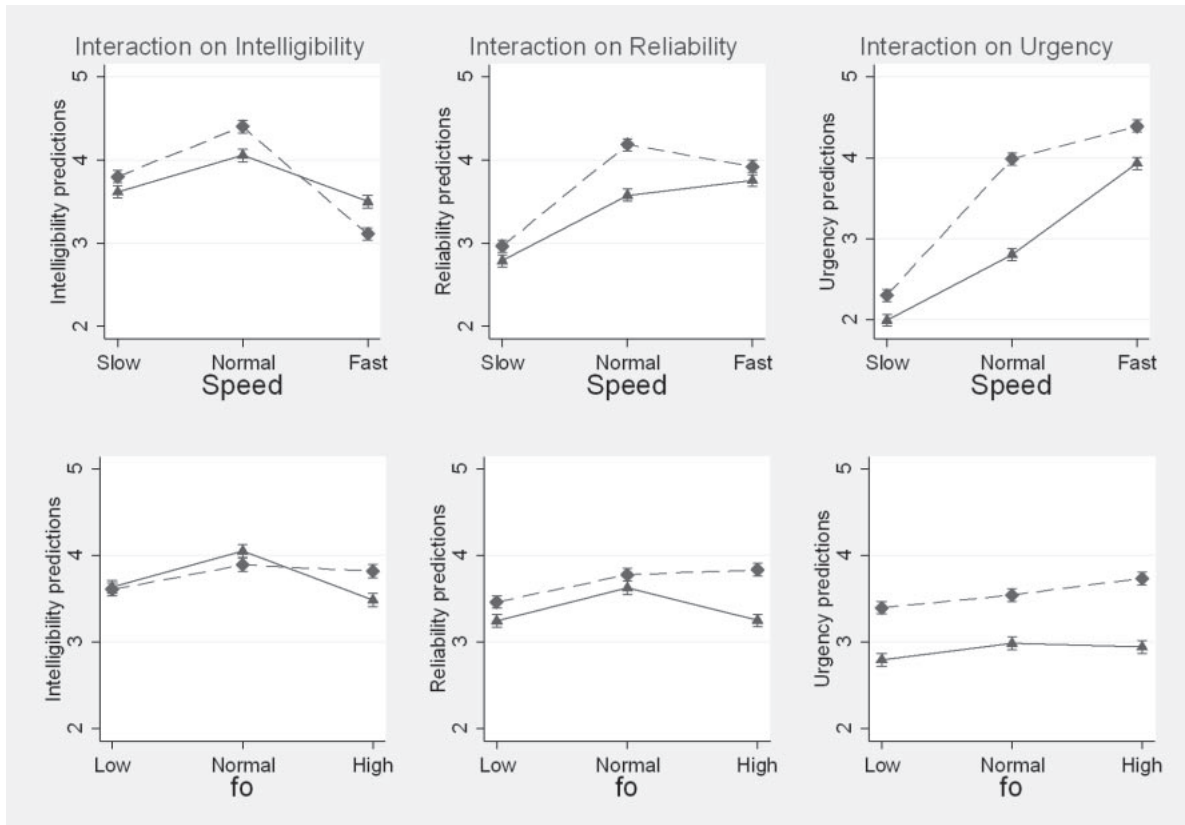


Fig. 3 Interaction plots for (a) Speed \times Speaker gender (top row), and (b) $f_0 \times$ Speaker gender (bottom row). From left to right, each two panels are for Intelligibility, Reliability and Urgency, respectively. Speaker gender key = \blacktriangle : male, \blacklozenge : female.

same column are for Intelligibility, Reliability, and Urgency, respectively. The scale of the vertical axis of all panels are the values on the choice scale of 1–5 on the questionnaire sheet used in the experiment, to mean 5 being the highest (most favorable) and 1 being the worst (least favorable) evaluation.

(a) Interaction Speaker gender \times Speed (Top three panels in Fig. 3)

All three graphs display that the two plots are not in parallel between the male- and the female-voice cases, implying the existence of interaction effects between Speaker gender and Speed. We interpret them that the influences from Speed on all three evaluations are different between male and female voices. Among them, when we look back to Table 2 we confirm that its effect size on Urgency was highest (0.23). The top right panel on Fig. 3 gives a graphical interpretation of that fact, where Speed gets faster, male voice increases urgency more markedly than female, whereas Speed gets slower, female voice sees a greater decrease in urgency than male voice.

(b) Interaction $f_0 \times$ Speaker gender (Bottom three panels in Fig. 3)

Similarly to (a), the two plots in each panel are not in parallel between male and female voices, again implying

the existence of that interaction effects although somewhat less conspicuous than the previous case (a). Indeed, the effect sizes and the associated F -values are all somewhat weaker than those of Speaker gender \times Speed on Table 2. The graphical commonality among the bottom three panels of Fig. 3 is that in male voice, the change in f_0 from normal to high do not result in a marked increase in the three evaluations, a clear contrast from the female voice cases. However, it is worth noting that, as we discussed in the main effect section, this result may be due in part to the fact that we have set $\pm 20\%$ for the variation in Speed, and up to ± 36 Hz for that in f_0 . Had we chosen a greater range of variation in f_0 , for example, the interaction Speaker gender \times f_0 may have been different; had it been greater, it may have been greater than the Speaker gender \times Speed interactions. Nevertheless, we believe that, under our choice of variation range that are reasonable in practical applications of evacuation calls, we are likely to experience this level of variation in both f_0 and Speed, and hence the degree of interaction effects observed this time.

A final note should be added, about the limits to the generalizability of our results. In Table 2, the control variables that govern inter-subject variability, sentence

types, venues of the experiment and the subject's age, are found to be statistically significant in particular in the Intelligibility and the Urgency evaluations. This implies the average evaluation scores may differ depending on these control characteristics.

4. DISCUSSION

Our results are both compatible and incompatible with the findings from the previous studies. Like other studies [15,17,18], we found that female voice was perceived better than male voice in terms of not only Urgency but also Intelligibility and Reliability, and that higher f_0 aroused the listeners a higher level of Urgency. The superiority of female voice for Intelligibility and Reliability can be derived from 'clarity' of voice characteristics in general. In his speech intelligibility test, Kwon [21] found that female speech obtained a significantly higher score than male speech, and his acoustic analyses showed that female speech had a higher f_0 and a wider vowel space, which suggests vowels are less centralized and reduced than in male speech. These gender characteristics are in accordance with the global characteristics of 'clear' speech summarized by Uchanski [22] and the higher f_0 of our female speaker shown in Table 1.

On the other hand, this study also revealed that normal f_0 obtained a higher rating than higher f_0 especially for Intelligibility. This suggests that the degree of raising f_0 for arousing perceived urgency should be carefully controlled when designing a verbal warning because it might sacrifice the level of intelligibility. Considering that one of the main functions of verbal warning is delivering disaster information, the decrease of intelligibility to a great extent by f_0 rising may not be desirable.

Another notable finding from this study is that as we had expected, speaking speed exhibited a greater effect on the three perception criteria (especially on perceived Urgency) compared with the gender or f_0 factor. As Fig. 3 illustrates, for Intelligibility, normal speed and even slower speed obtained higher rating than faster speed. Apparently, the faster a warning is spoken, the harder it is to understand what is spoken. On the other hand, it is worthy to note that for Reliability, there is only a slight gap in rating between normal speed and faster speed. This relatively high Reliability for a faster spoken warning could be explained in relation with perceived urgency referring to Robinson's psychological model [23]. The model illustrates that urgency is detected unconsciously at a very low level of psychological processing to identify a stimulus, and it results in automatic construction of fear, which makes the receiver of the stimulus quickly respond to it physically and psychologically. In our perception experiment, faster speed raised the Urgency level to the highest. It is possible that the listeners automatically judged

Table 3 Correlations between perception criteria.

	Intelligibility	Reliability	Urgency
Intelligibility	1.00	—	—
Reliability	0.38	1.00	—
Urgency	0.10	0.65	1.00

a faster spoken warning as reliable as a quick reaction to the stimuli without taking enough time to doubt or at least consider its reliability due to the rapid construction of urge and fear even in the experimental setting. The relatively high correlation ($r = 0.65$) between Urgency and Reliability in Table 3 could support this hypothesis. Likewise, the previous perception studies using sound alerts [12,13] found shorter- or non-interval alerts exhibited better perception as an alert, which can be explained in the same fashion: the ceaseless sound stimulated automatic arousal of fear and listeners reacted quickly and unconsciously. Care must be taken to compare the results of these preceding studies with ours, though, in terms of the magnitude of variations chosen for f_0 and speed, as these parameter choice may affect the degree of influences observed in respective experiments.

As mentioned earlier, disaster warnings should be efficient and effective to overcome normalcy bias in order to encourage evacuation behaviors. People tend to underestimate the possibility and risk of a disaster based on their own knowledge, past experience, and other social elements such as neighbor influence. This cognitive stage seems to be at the conscious level, which comes after the unconscious level, where automatic arousal of fear occurs. Based on the results from our study, we argue that the speech rate may effectively be varied depending on the purpose of an evacuation call, whether it prioritizes urgency, or intelligibility and reliability. In order to heighten perceived urgency and reliability, the words intended to attract listeners' attention (e.g., "Emergency alert!" "Evacuate immediately!") can be pronounced with a higher f_0 and faster speed. On the other hand, crucial disaster information, such as expected tsunami arrival time and height, the place of evacuation shelters, roads blocked and roundabouts, should be spoken with a normal f_0 and speed to maximize intelligibility. It may not be easy for a caller to vary speed and f_0 during speech; hence, those who are in charge of articulating a warning need to have enough training and practice to maintain a good control of these acoustic features in their speech.

5. CONCLUSIONS

In this research, three parameters (voice gender, speed, and f_0) were tested in order to find out the best combination of factors that enhances perceived intelligibility, reliability,

and urgency when the listeners hear the warnings.

Below are our main findings:

- When the speaker gender is female, the average evaluations for Reliability and Urgency have improved, than when it is male.
- For both speaking speed and f_0 , setting them to normal (compared from slow/fast (+/−20%) for speed, and from low/high (+/− up to 36 Hz) for f_0) improved the average evaluations for Intelligibility and Reliability. For Urgency only, setting speed to faster (both slow to normal and normal to fast) or setting f_0 to higher (both low to normal and normal to high) resulted in an improved average evaluation.
- For all of Intelligibility, Reliability, and Urgency, the main effect of speaking speed was the most dominant. In particular, Urgency can be influenced by the speed factor alone by up to 39%. By setting speed to fast (+20%), all other things being equal, the average perceived Urgency raised to 4.0 on the 1–5 scale, from 3.2 when the speed is set to normal. This can be effective even at the expense of Intelligibility and Reliability to some degrees.
- Based on these results, we argue that the speech rate may effectively be varied depending on the purpose of an evacuation call, whether it prioritizes urgency, or intelligibility and reliability.

There are some issues to be solved in the future study:

1) we wish to use similar experiments to test other respondents who reflect actual demographic traits (age in particular) of residents in the disaster-prone areas, such as coastal areas threatened by tsunami and riverside basins threatened by floods; 2) the auditory condition may have to incorporate actual environmental limitations experienced by such dwellers, where, for example, evacuation calls can be difficult to hear in torrential rain and high winds; 3) the content, in addition to the auditory properties examined this time, of the evacuation calls deserves to be studied as it will also significantly affect respondents' perception.

ACKNOWLEDGMENTS

This work was supported by JSPS KAKENHI Grant Number JP26370490 and the Foundation of River & Basin Integrated Communications, JAPAN grant-in-aid #27-8. We gratefully acknowledge the assistance of Akari Harada and all the participants in our experiment.

REFERENCES

- [1] T. Sumi, "7-2 Kikiyosoku to hinan (Prognosis for danger)," in *Saigaijohogaku Jiten (Dictionary for Disaster Information)*, Nihon Saigaijoho Gakkai, Ed. (Asakura Shoten, Tokyo, 2016), pp. 184–185 (in Japanese).
- [2] I. Nakamura, "Hinan no riron (Evacuation Theory)," in *Saigai Kikikanriron Nyuumon (Introduction to Disaster Risk Management)*, H. Yoshii and A. Tanaka, Eds. (Koubundou, Tokyo, 2008), pp. 154–163 (in Japanese).
- [3] J. H. Sorensen and B. V. Sorensen, "Community process: Warning and evacuation," in *Handbook of Disaster Research*, H. Rodriguez, E. L. Quarantelli and R. R. Dynes, Eds. (Springer Science+Business Media, LLC., New York, 2007), pp. 183–199.
- [4] T. E. Drabek, "Understanding disaster warning responses," *Soc. Sci. J.*, **36**, 515–523 (1999).
- [5] E. D. Kuligowski and S. M. Gwynne, "The need for behavioral theory in evacuation modeling," *Pedestrian Evacuation Dynamics 2008*, February 28–29, 2008, Wuppertal, Germany (2008).
- [6] C. B. Mayhorn and A. C. McLaughlin, "Warning the world of extreme events: A global perspective on risk communication for natural and technological disaster," *Saf. Sci.*, **61**, 43–50 (2014).
- [7] W. Donner, H. Rodriguez and W. Diaz, "Public warning response following tornadoes in New Orleans, LA, and Springfield, MO: A sociological analysis," *Second Symposium on Policy and Socio-economic Research*, Jan 15–18, 2007, San Antonio, TX, USA (2007).
- [8] T. Katada, M. Kodama, N. Kuwasawa and S. Koshimura, "Juumin no hinankoudou ni miru tsunami bousai no genjou to kadai: 2003nen Miyagiken-oki no jishin · Kesennuma shimin ishikichousa kara (Issues of resident's consciousness and evacuation from the tsunami: From questionnaire survey in Kesennuma city, Miyagi Pref. after the earthquake of Miyagiken-oki, 2003)," *Dobokugakkai Ronbunshuu (J. Jpn. Soc. Civ. Eng.)*, **789** (II-71), 93–104 (2005) (in Japanese).
- [9] H. Inoue, "Ooaraimachi wa naze hinan seyo to yobikaketanoka: Higashinihondaishinsai de bousai gyousei musen housou ni tsukawareta yobikake hyougen no jirei houkoku (Why did Oarai town use the imperative form in alerting the residents to evacuate?: A report on phrases used in the disaster prevention administrative radio system when the Great East Japan Earthquake hit)," *Housou Kenkyuu to Chousa (NHK Mon. Rep. Broadcast Res.)*, Sep., pp. 32–53 (2011) (in Japanese).
- [10] K. Arai, "How to transmit disaster information effectively: A linguistic perspective on Japan's tsunami warnings and evacuation instructions," *Int. J. Disaster Risk Sci.*, **4**, 150–158 (2013).
- [11] I. R. Murray and J. L. Arnott, "Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion," *J. Acoust. Soc. Am.*, **93**, 1097–1108 (1993).
- [12] M. S. Wogalter, V. C. Conzola and T. L. Smith-Jackson, "Research-based guidelines for warning design and evaluation," *Appl. Ergon.*, **33**, 219–230 (2002).
- [13] E. C. Haas and J. Edworthy, "Designing urgency into auditory warnings using pitch, speed and loudness," *Comput. Control Eng. J.*, **7**, 193–198 (1996).
- [14] C. Suied, P. Susini and S. McAdams, "Evaluating warning sound urgency with reaction times," *J. Exp. Psychol.: Appl.*, **14**, 201–212 (2008).
- [15] J. Edworthy and E. Hellier, "Urgency in speech warnings," *Proc. 8th Int. Congr. Noise as a Public Health Problem*, T. Houtgast, E. A. M. Fransten and W. F. Hofman, Eds. (Foundation ICBEN, Rotterdam, 2003), pp. 65–69.
- [16] K. S. Park and P. S. Jang, "Effects of synthesized voice warning parameters on perceived urgency," *Int. J. Occup. Saf. Ergon.*, **5**, 73–95 (1999).
- [17] E. Hellier, J. Edworthy, B. Weedon, K. Walters and A. Adams, "The perceived urgency of speech warnings: Semantics versus acoustics," *Hum. Factors*, **44**, 1–17 (2002).

- [18] S. Machado, E. Duarte, J. Teles, L. Reis and F. Rebelo, "Selection of a voice for a speech signal for personalized warnings: The effect of speaker's gender and voice pitch," *Work*, **41**, Suppl. 1, 3592–3598 (2012).
- [19] P. Boersma and D. Weenink, "Praat: Doing phonetics by computer," 2017. [Computer program Version 6.0.26]. Available: <http://www.praat.org/> (Accessed: 2-March-2017).
- [20] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. (Lawrence Erlbaum Associates, Hillsdale, N.J., 1988).
- [21] H.-B. Kwon, "Gender difference in speech intelligibility using speech intelligibility tests and acoustic analyses," *J. Adv. Prosthodont.*, **71**(6), pp. 71–76 (2010).
- [22] R. M. Uchanski, "Clear speech," in *The Handbook of Speech Perception*, D. B. Pisoni and R. E. Remez, Eds. (Blackwell Publishing Ltd., Malden, MA, 2005), pp. 207–235.
- [23] M. D. Robinson, "Running from William James' bear: A review of preattentive mechanisms and their contributions to emotional experience," *Cogn. Emot.*, **12**, 667–696 (1998).

APPENDIX

The effect size values used throughout this paper are

the ANOVA f effect size, that is calculated using the partial correlation ratio η_p^2 for factor A using the below formula:

$$\eta_p^2 = (d.f._A \times F_A) / (d.f._A \times F_A + d.f._{error}) \quad (A.1)$$

where $d.f.$ is degree of freedom and F is F -value ($d.f._{error}$ is error's degree of freedom). Then sample effect size f is derived by:

$$f = \sqrt{\eta_p^2 / (1 - \eta_p^2)} \quad (A.2)$$

For example, in Table 2, the sample effect size of $f = 0.03$ of Main effects of Speaker gender for Intelligibility was derived as follows:

$$\eta_p^2 = 1 \times 2.3 / (1 \times 2.3 + 2,926) = 7.9 \times 10^{-4}$$

$$f = \sqrt{\eta_p^2 / (1 - \eta_p^2)} = 0.028 \approx 0.03$$

In this paper, the floor values for classifying the effect size are chosen to be 0.1 as "small," 0.25 as "medium," and 0.4 for "large," taking note of Cohen (1988) [20].