*INTERSPEECH 2016*
September 8–12, 2016, San Francisco, USA

# Effects of urgent speech and preceding sounds on speech intelligibility in noisy and reverberant environments

*Nao Hodoshima*

Department of Information Media Technology, Tokai University, Japan
hodoshima@tokai-u.jp

## Abstract

Public-address (PA) announcements are used to convey emergency information; however, noise and reverberation sometimes make announcements in public spaces unintelligible. Therefore, the present study investigated how combinations of speech spoken in an urgent style and preceding sounds affect speech intelligibility and perceived urgency in noisy and reverberant environments. Sentences were spoken in normal or urgent styles and preceded by either two sounds (siren sound or ocean wave-like sound) or no sounds. Eighteen young participants carried out word identification test and rated perceived urgency on five-point scales in noisy and reverberant environments. The results showed that the urgently spoken speech had significantly higher speech intelligibility than the normal speech. The urgently spoken speech preceded by the wave-like sound showed significantly higher speech intelligibility than normal speech without sounds, normal speech preceded by the siren sound, and urgently spoken speech preceded by the siren sound. The results also demonstrated that the perceived urgency was rated higher for the urgently spoken speech than that for the normal speech, regardless of the types of preceding sounds. These results suggest that appropriate combinations of speaking styles and alerting sounds will increase the intelligibility of emergency PA announcements.

**Index Terms**: speech intelligibility, urgency, reverberation, noise

## 1. Introduction

Public address (PA) announcements are sometimes difficult to interpret due to noise and reverberation. For example, there exists an average ambient noise level of 82 $dBL_{Aeq,5h}$ at 19 railway station platforms in Japan [1] and reverberation times were longer than 2 s for 500-2k Hz octave bands in two train stations in Tokyo [2]. Speech intelligibility in noisy and/or reverberant environments is generally lower for older adults and non-native listeners compared with young native listeners [3, 4]. Since the population is rapidly aging (e.g., as of 2015 in Japan, the population of people aged ≥65 was 26% of the total population [5]), care must be taken when broadcasting announcements in public spaces, especially in emergency situations (e.g., fire or earthquake).

People can modify the way they speak, therefore affecting their speech intelligibility, according to their current situation and surrounding acoustical environments. Clearly articulated speech (clear speech) has higher word intelligibility than conversational speech for people with and without hearing impairments in quiet, noise, and reverberant environments [6-8]. Acoustical analysis showed that vowel duration was longer and vowel space was greater in clear speech than those in conversational speech [9].

The Lombard effect (when individuals increase their vocal effort in the presence of noise) [10] is another example of speech modification. Speech spoken in noisy environments yielded higher word intensity, duration, fundamental frequency, and first formant frequency as well as higher word identification scores than quiet environments when heard in a noisy environment [10-13]. When individuals speak in the presence of reverberation (reverberation-induced speech), the acoustic characteristics and speech intelligibility of that speech are increased similarly to those observed in the Lombard effect, despite the masking patterns of the noise and reverberations being temporally and spectrally different [13-15].

Speaking slowly is another way to increase speech intelligibility, especially in reverberant environments because it reduces overlap-masking (i.e., reverberant phonemes mask the following ones) [16]. As an example of this, when older adults speak at an increased rate in noisy or reverberant environments, lower speech intelligibility is observed compared to younger adults [17]. Slowed speaking rate with a time-delayed technique, in an environment in which speech sounds broadcasted from loudspeakers are delayed by the distance between adjacent loudspeakers and the speed of sound, decreased perceived listening difficulty of young adults compared to normal-speed speech without the time-delay technique [18]. However, speaking too slowly is not appropriate for emergency announcements as it might be difficult to perceive the urgency.

The perceived urgency (subjective rating on urgency that is interpreted from a sound) of a complex pulse was increasingly rated higher as its fundamental frequency and sound pressure level increased [19]. Speech spoken in an urgent style had higher fundamental frequency, broader fundamental frequency range, and higher amplitude yielded higher perceived urgency than speech spoken in a normal style [20]. Urgently spoken words yielded faster response times compared with normally spoken words [21]. Since most studies focused on what kind of characteristics of non-speech and speech sounds affect subjective ratings of urgency, very few studies have investigated the relationship between speech spoken urgently and speech intelligibility.

The goal of this study was to make PA announcements more intelligible in noisy and reverberant environments by modifying broadcasting speech itself rather than implementing architectural acoustical and/or electroacoustical solutions. The current study investigated whether speech spoken in an urgent style improved speech intelligibility compared with speech spoken in a normal style in noisy and reverberant environments. This study also investigated whether preceding sounds to urgent speech affect speech intelligibility. Since a



http://dx.doi.org/10.21437/Interspeech.2016-1618

typical urgency sound, such as a siren, and urgent speech both increase perceived urgency, the second purpose of this study was to examine whether combination of both an urgent sound and urgent speech would further increase speech intelligibility in noisy and reverberant environments.

## 2. Listening test

### 2.1. Participants

The participants were 18 native speakers of Japanese (15 males, 3 females; average age, 20 years). All had self-reported normal hearing.

### 2.2. Stimuli

Speech materials consisted of a target word embedded in the carrier sentence "A fire has broken out. Evacuate to (Target)". A total of 60 target words of four morae (a phonological syllable-like unit in Japanese) were selected from a database of familiarity-controlled Japanese word lists (FW07) [22]. Word familiarity in this study was chosen between 1.0 and 2.5 on a seven-point scale (1 for the least familiar and 7 for the most familiar) [22] in order to avoid the participants using context cues as well as semantic cues of the target words. While sematic information of words changed the perceived urgency rate [20], this study focused on bottom-up cues in order to investigate the direct effect of speech spoken in an urgent style on speech intelligibility.

The speaker was a 24-year-old female native speaker of Japanese. She reported neither hearing nor speaking disorders. She had received voice training at a voice acting school in Tokyo for 2 years and had experience with an amateur voice cast in a play. Since the present study was a preliminary study, a single speaker was chosen rather than several speakers who might have wider variations in speech production.

Two speaking conditions were used in this study: urgent and normal. In the urgent condition, the speaker was instructed to imagine that she was going to make PA announcements that a fire had broken out in a train station and to warn passengers with urgency about the emergency. In the normal condition, the speaker was instructed to speak as she speaks normally in conversation.

Speech sounds were recorded on a personal computer through a microphone (SHURE KSM141; condenser, cardioid) and a digital audio interface (TASCAM US-144MKII) in a sound-treated room. After speech sounds were recorded, one carrier sentence was chosen for each speaking condition, and the target words with 100 ms preceding and following pauses were embedded in the carrier sentence for each speaking condition. This was done in order to control the effect of overlap-masking on the target words. The intensity ratio of the carrier sentence relative to the target word was normalized to the speaking condition.

Three sound conditions were used in this study: no sound, a siren sound, and an ocean wave-like sound preceding speech. The siren sound was used to simulate urgency in an emergency situation as well as to gain attention, while the ocean wave-like sound was used to evoke relaxation. The intention of using these conditions was to study whether speech intelligibility increases 1) by alerting and providing a sense of urgency with a preceding siren sound; and/or 2) by evoking relaxation with the preceding ocean wave-like sound and thereby subconsciously raising awareness so that people will concentrate more on what they are hearing. The siren was repetitions of a linearly increasing swept sine wave from 770 to 960 Hz over 300 ms and a linearly decreasing swept sine wave from 960 Hz to 770 Hz over 300 ms with total duration of 5.4 s. The frequency and duration of the swept sine wave were intended to mimic widely-used alarm sounds of an ambulance in Japan. The ocean wave-like sound was made by rotating red beans on a 60 cm-diameter tilted and inverted umbrella made of vinyl; it was recorded on a personal computer through the microphone and the digital audio interface in the sound-treated room. The recorded sound was passed through a finite impulse response filter using Adobe Audition software, which reduced gains at peak frequency from 100 to 1k Hz and from 8k to 20k Hz by up to 6 dB; boosted gains at peak frequency from 1k to 4k Hz by up to 3 dB; applied a pitch-shifter to shift pitch by -55 cents; and convolved the sound with an impulse response with a Decay time of 500 ms, Pre-decay time of 200 ms, Diffusion of 50 ms. The sound was then passed through a de-emphasis filter in which the spectral slope decreased by 6 dB/octave above 1500 Hz using Praat software. The filters were applied in order to produce a recorded sound as close to a real ocean wave sound as possible temporally and spectrally. The ocean wave-like sound was fabricated as recording real ocean wave sounds often includes background noises at the seashore near the Tokyo area (e.g., cars, bikes, trains, people talking/yelling), which are difficult to remove afterwards, even at midnight; in addition, available copyright-free CDs of recorded ocean wave sounds contained extra sounds (e.g., keyboard sounds). The recorded sound was determined to be an ocean wave sound by three people who did not know the research.

Table 1. *Experimental conditions.*

| Condition | Details |
|---|---|
| 1 | Normal speech |
| 2 | Urgent speech |
| 3 | Ocean wave-like sound + normal speech |
| 4 | Ocean wave-like sound + urgent speech |
| 5 | Siren + normal speech |
| 6 | Siren + urgent speech |

The intensity ratio of the sounds relative to the speech sounds was normalized, and the sounds were inserted preceding the speech sounds with 50 ms pauses, making the six experimental conditions shown in Table 1.

All stimuli were added with a babble noise (a mixture of four utterances of two male speakers from a speech database [23] with a signal to noise ratio of 10 dB) and then convolved with an impulse response (reverberation time of 2.0 s for octave bands from 125 to 4000 Hz) using Matlab software to simulate an average listening environment at a train station installed with sound-reflective walls. The overall intensity of the stimuli was normalized across the conditions. The total number of stimuli was 362 (6 experimental conditions × 60 sentences + 2 sentences used for a practice session).

### 2.3. Procedures

The listening test was performed in a sound-treated room where stimuli were presented to each participant diotically through headphones (STAX SR-303; electrostatic, open circumaural type) through a digital audio interface (TASCAM US-144MKII) connected to a computer. Two practice trials

were held in order to familiarize the participants with the experimental procedure. The playback level was adjusted to each participant's comfort level.

In each trial, a stimulus was presented once, and the participants were instructed to write down what they heard as a target word on their answer sheets. The participants then were asked to rate the impression of the stimulus relating to perceived urgency on 5-point scales. The rating scales used five adjective pairs relating to perceived urgency and comprised strong-weak, pleasant-unpleasant, powerful-not powerful, slow-quick, and safe-dangerous pairs (1 corresponds to the first word in each pair, 5 corresponds to the second word in each pair). The five adjective pairs were based on the 16 adjective pairs which had been used in the previous study on dangerousness of evaluating sounds [24]. Instead of asking participants about perceived urgency directly as in the previous studies [19-21], this study used the adjective pairs to investigate how the preceding sound and urgent speech separately affect participants' impressions in detail. For each participant, 60 stimuli (6 conditions × 10 sentences) were presented randomly. The target word and condition combinations were randomized across the participants.

## 3. Results and discussion

Figure 1 shows the mean percentage of correct mora identification rate of target words. Statistical analyses were carried out using IBM SPSS Statistics, and the significance level was set at 5% in the present study. A 2 3 ANOVA was carried out with the speaking condition (normal and urgent) and the preceding sound (no sound, ocean wave-like sound and siren) as repeated variables and the correct mora identification rate of target words as the dependent variable.

The main effect of speaking condition was significant ($F(1,17)=7.213$, $p=0.016$), indicating that urgent speech was significantly more intelligible than normal speech. The main effect of the preceding sound was significant ($F(2,34)=4.905$, $p=0.041$), and a post-hoc test revealed significant differences between no sound and the ocean wave-like sound, and between the ocean wave-like sound and the siren, indicating that correct mora identification rate of target words preceded by the ocean wave-like sound was significantly higher than that of speech either preceded by the siren or without sound. Multiple comparisons showed that urgent speech preceded by the wave-like sound showed significantly higher speech intelligibility than normal speech without sounds ($p=0.012$), normal speech preceded by the siren sound ($p=0.049$), and urgent speech preceded by the siren sound ($p=0.009$).

Figure 2 shows the mean ratings of impressions of the stimuli. The speaking style rather than preceding sound affected perceived urgency (strong, powerful, quick, and dangerous), although there was no significant difference between experimental conditions based on the Friedman rank test.

As expected, urgent speech was significantly more intelligible than normal speech, and urgent speech was rated stronger, more powerful, quicker, and more dangerous than normal speech, regardless of the preceding sound. These results indicate that urgent speech not only increased perceived urgency, as reported in previous studies [19-21], but also improved speech intelligibility in noisy and reverberant environments.

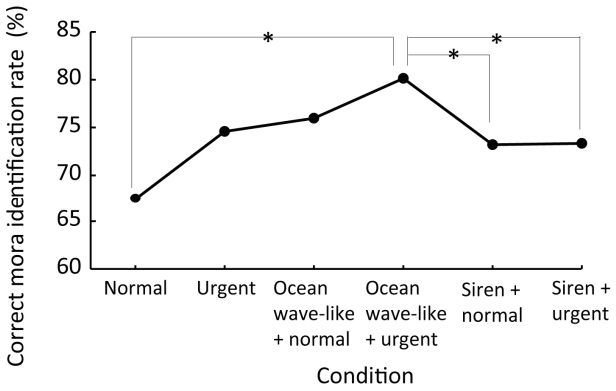


Figure 1: *Mean correct mora identification rate of target words in speech spoken normally or urgently preceded by a siren sound, an ocean wave-like sound, or no sound. Asterisks indicate a significant difference between conditions ($p < 0.05$).*

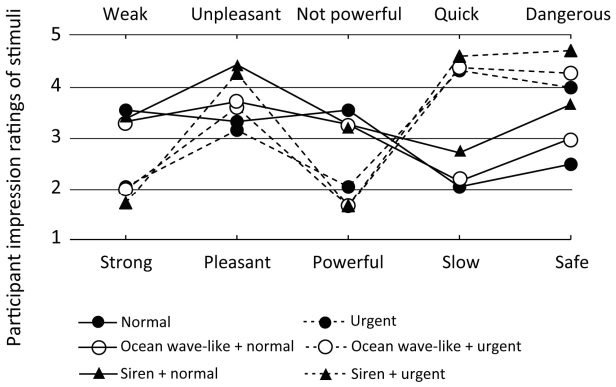


Figure 2: *Mean participant impression ratings of speech spoken normally or urgently preceded by a siren sound, an ocean wave-like sound, or no sound.*

Moreover, urgent speech preceded by the wave-like sound showed significantly higher speech intelligibility than normal speech without sounds, normal speech preceded by the siren sound, and urgent speech preceded by the siren sound. This indicates that not only speaking style, but also the type or lack of preceding sound affected speech intelligibility. The combination of both urgent conditions, the siren sound and urgent speech, did not further increase speech intelligibility, whereas the ocean wave-like sound followed by urgent speech did increase intelligibility. These results imply that the ocean wave-like sound unconsciously raised listener attention to the incongruent emergency PA announcements so that participants would concentrate more on the target words. However, the evidence for this is not clear in the limited conditions of this study, and further research is needed on which combination of preceding sound and urgent speech increases speech intelligibility.

Acoustical analysis revealed that fundamental frequency was higher in urgent speech than normal speech (the average increase was 35.8 Hz for the target words embedded in the carrier sentence and 28.8 Hz for the target words respectively), which is consistent with the previous studies [19-21]. Since clear speech [9], the Lombard speech [11], and the reverberation-induced speech [13], which improved speech

intelligibility in noisy and/or reverberant environments were reported to increase fundamental frequency compared with normally spoken speech [9, 11, 13], fundamental frequency may be one of the contributors that improves speech intelligibility in noisy and/or reverberant environments. The higher impression of perceived urgency (stronger, more powerful, quicker, and more dangerous) of urgent speech than normal speech may reflect higher fundamental frequency in urgent speech. Other acoustical parameters such as word duration, formant frequencies, and an average spectral tilt should be further analyzed. These are not described in the previous perceived urgency [19-21] but in the previous studies on clear speech, Lombard speech, reverberation-induced speech [8-15].

This study used a single female speaker, as female speakers tend to yield higher ratings of perceived urgency than male speakers [21]. For clear speech, female speakers were also shown to be more intelligible than male speakers [9]. Future research on whether female speakers are more intelligible for urgent speech than male speakers is needed.

## 4. Conclusions

The present study investigated how combinations of speaking styles (normal and urgent) and preceding sounds (no sound, a siren sound, or an ocean wave-like sound) affect intelligibility of words in a sentence and perceived urgency in noisy and reverberant environments. The results showed that urgent speech was significantly more intelligible than normal speech. Urgent speech preceded by the wave-like sound showed significantly higher speech intelligibility than normal speech without sounds, normal speech preceded by the siren sound, and urgently spoken speech preceded by the siren sound. The results also demonstrated that the perceived urgency (strong, powerful, quick, and dangerous in this study) was rated higher for urgent speech than that for normal speech, although there was no significant difference, regardless of the types of preceding sounds, which agrees with previously reported studies on perceived urgency. Since the wave-like sound followed by urgent speech was more intelligible but rated less dangerous than normal speech, it will be interesting to study how much the wave-like sound yields an "alerting effect" which would be required in emergency situations. As this preliminary study used one female speaker and limited preceding sounds, future research with an increased number of speakers and study combinations of urgent speech and preceding sounds is needed. It would be desirable to expand this study to include diverse listeners such as older adults and non-native listeners, and also to test intelligibility for speech presented through loudspeakers that is similar to those used in public spaces. Appropriate combinations of speaking styles and alerting sounds will further increase intelligibility of emergency PA announcements in public spaces.

## 5. Acknowledgements

The author is grateful to the speaker and listeners who participated in this study, and to Katsuo Morio of Tokai University for conducting the listening tests.

## 6. References

[1] R. Shimokura, and Y. Soeta, "Characteristics of train noise in above-ground and underground stations with side and island platforms", *J. Sound and Vibration*, vol. 330, pp. 1621–1633, 2011.

[2] Y. Izumi, "Actual condition of acoustical environment in railway station", *J. Aoust. Soc. Jp*., vol. 70, no. 3, pp. 116-122, 2014. (in Japanese)

[3] A. K. Nablek and P. K. Robinson, "Monaural and binaural speech perception in reverberation for listeners of various ages", *J. Acoust. Soc. Am*., Vol. 71, pp. 1242-1248, 1982.

[4] A. K. Nabelek and A. M. Donahue, "Perception of consonants in reverberation by native and non-native listeners", *J. Acoust. Soc. Am*., vol. 75, no. 2, pp. 632-634, 1984.

[5] The Cabinet Office, "Annual Report on the Aging Society", Japan, 2015.

[6] M. A. Picheny, N. L. Durlach, and L. D. Briada, "Speaking clearly for the hard of hearing I", *J. Speech Hear. Res*., vol. 28, pp. 96-103, 1985.

[7] K. L. Payton, R. M. Uchanski, and L. D. Braida, "Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing", *J. Acoust. Soc. Am*., vol. 95, no. 3, pp. 1581-1592, 1994.

[8] R. Caissie, M. M. Campbell, W. L. Frenette, L. Scott, I. Howell, and A. Roy, "Clear speech for adults with a hearing loss: Does intervention with communication partners make a difference", *J. Am. Acad. Audiol.*, vol. 16, pp. 157-171, 2005.

[9] S. H. Ferguson and D. Kewley-Port, "Talker differences in clear and conversational speech: Acoustic characteristics of vowels", *J. Speech. Lang. Hear. Res*., vol. 50, pp. 1241-1255, 2007.

[10] H. Lane and B. Tranel, "The Lombard sign and the role of hearing in speech", *J. Speech Hear. Res.*, vol. 14, pp. 677-709, 1971.

[11] W. Van Summers, D. B. Pisoni, R. H. Bernacki, R. I. Pedlow and M. A. Stokes, "Effects of noise on speech production: Acoustics and perceptual analysis", *J. Acoust. Soc. Am*., vol. 84, pp. 917-928, 1988.

[12] J. C. Junqua, "The Lombard reflex and its role on human listeners and automatic speech recognizers", *J. Acoust. Soc. Am*., vol. 93, pp. 510-524, 1993.

[13] N. Hodoshima, T. Arai and K. Kurisu, "Speaker variabilities of speech in noise and reverberation", *IEICE Technical Report,* SP2009-69, pp. 43-48, 2009. (in Japanese)

[14] N. Hodoshima, T. Arai and K. Kurisu, "Intelligibility of speech spoken in noise and reverberation", *Proc. International Congress on Acoustics* (paper ID: 663), 2010.

[15] N. Hodoshima, T. Arai, and K. Kurisu, "Intelligibility of speech spoken in noise/reverberation for older adults in reverberant environments", *Proc. Interspeech* (paper ID: P6a.06), 2012.

[16] A. K. Nabelek, T. R. Letowski and F. M. Tucker, "Reverberant overlap- and self-masking in consonant identification", *J. Acoust. Soc. Am*., vol. 86, pp. 1259-1265, 1989.

[17] S. Gordon-Salant and P. J. Fitzgibbons, "Recognition of multiply degraded speech by young and elderly listeners", *J. Speech Hear. Res*., vol. 38, pp. 1150-1156, 1995.

[18] S. Yokoyama, S. Sakamoto, H. Tachibana and S. Tazawa, "Study on the application of time-delay technique to public address system in a tunnel", *Proc. Inter-noise*, 2005.

[19] E. C. Haas and J. Edworthy, "Designing urgency into auditory warnings using pitch, speed and loudness", *J. Computing & Control Engineering*, vol. 7, no. 4, pp. 193-198, 1996.

[20] E. Hellier, J. Edworthy, B. Weedon, K. Walters, A. Adams, "The perceived urgency of speech warnings: Semantics versus acoustics" *J. Human Factors*, vol. 44, no. 1, pp. 1-17, 2002.

[21] J. K. Ljungberg and F. Parmentier, "The Impact of intonation and valence on objective and subjective attention capture by auditory alarms", *J. Human Factors*, vol. 54, no. 5, pp. 826-37, 2012.

[22] S. Amano, T. Kondo, S. Sakamoto and Y. Suzuki, "Familiarity-controlled word lists 2003 (FW03)", The Speech Resources Consortium, National Institute of Informatics in Japan, 2006.

[23] "Phonetically-balanced 1000 sentences speech database", NTT Advanced Technology Corporation, 1999.

[24] S. Kuwano, S. Namba, A. Schick, H. Höge, H.Fastl, T. Filippou and M. Florentine, "Subjective impression of auditory danger signals in different countries", *Acoust. Sci. and Tech.*, vol.28, no. 5, pp. 360-362, 2007.