Mathematical Modeling of the Effects of Speech Warning Characteristics on Human Performance and Its Application in Transportation Cyberphysical Systems

Yiqi Zhang, Changxu Wu, *Member, IEEE*, and Jingyan Wan

***Abstract*—Transportation cyberphysical systems (CPSs) aim to improve driving safety by informing drivers of hazards with warnings in advance. The understanding of human responses to speech warnings is essential in the design of transportation CPSs to eliminate hazards and accidents. To date, many works have addressed diverse warning characteristics with experimental approaches. However, the computational model to quantify the effects of warning characteristics on human performance in responses to speech warnings is still missing. Mathematical equations were built to model the effects of lead time, loudness, and signal word choices on human perceptual, cognitive, and motor activities involved in speech warning responses. Different levels of lead time, levels of loudness, and signal word choices served as inputs in the model to predict human error rate and reaction time of speech warning responses. The model was validated with drivers' crash rates and reaction times to speech warnings of upcoming hazards in driving assistant systems in two empirical studies. Results show a good prediction of human performance in responding to speech warnings compared with the empirical data. The application of the model to identify optimal parameter settings in the design of speech warnings in order to achieve greater safety benefits is later discussed.**

***Index Terms*—Human performance modeling, human–computer interaction, intelligent transportation systems.**

I. Introduction

DEATHS and injuries resulting from road traffic accidents has become a major public health problem. According to statistic data published by the National Highway Traffic Safety Administration (NHTSA) in U.S., 5.3 million crashes occurred nationally in 2011 [1]. With regard to improve driving safety, recent advances in Transportation Cyber-Physical Systems (CPS) aim to establish a connected transportation environment by monitoring the status of the physical worlds (e.g., sensors and actuators), connecting it with the cyber worlds (e.g., information, communication, and intelligence), and providing the integrated real-time information among multiple levels, including vehicles to vehicle communication, vehicle to infrastructures communication and in-vehicle information communication [2]. Compared to conventional transportation environment, the connectivity of the transportation CPS allows drivers to learn about the traffic status out of their sight, and provides them with more time to respond to warnings regarding potential hazards [3].

In order to improve the safety of both humans and vehicles, as well as facilitate communication between them, it is important to design warning characteristics based on human performance. While work has been done to increase the communication reliability of connected vehicles, the effectiveness of such systems could not be achieved without drivers making proper and timely responses. Therefore, modeling driver responses to warnings is necessary to achieve effectiveness of warning systems with the human in the loop.

Compared to non-speech auditory warnings, speech warnings are more user-friendly since humans can easily understand and differentiate warnings without specific trainings in memorizing and recognizing warnings [4]. Previous work showed that people working in an operation room had difficulties in recognizing more than half of the non-speech warnings currently in use [5]. Another study indicated that people were unable to distinguish more than six complex warnings [6]. Moreover, previous work found that speech warnings led to a faster reaction time than non-speech warnings regarding spatial information [7]. As a consequence, speech warnings can be widely applied to the Transportation CPS with different warnings in diverse traffic situations.

To date, many empirical studies have examined the influence of warning characteristics on human performance, such as content, perceived hazard, familiarity, signal word, warning sources, and number of items in speech warnings, on human behavior and performance [8]–[11]. Existing empirical has been shown that warning lead time, loudness and signal word choice have significant effects on driver responses to speech warnings. Lead time is defined as the available time for responses from the start of the speech warning until the occurrence of the collision [12]. Studies showed early warnings led to shorter reaction times to collisions than either middle or late warnings [13]–[15]. The warning loudness was found to have a significant effect on urgency expression [16]. In terms of warning

Manuscript received September 14, 2015; revised January 21, 2016; accepted February 27, 2016. This work was supported by the National Science Foundation. The Associate Editor for this paper was Y. Gao.

The authors are with the Department of Industrial and Systems Engineering, The State University of New York at Buffalo, Buffalo, NY 14260 USA (e-mail: yiqizhan@buffalo.edu; seanwu@buffalo.edu; jingyanw@buffalo.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2016.2539975

semantics, the different signal words chosen in speech warnings significantly influence a human's judgment of the urgency level of a situation [17]. However, the behavioral approach used in existing empirical studies to assess the effectiveness of speech warnings can be highly task-dependent, time consuming, and high-cost. The modeling approach we adopted in the current work will provide the predictions of human performance under the different levels of the modeled warning characteristics by running the developed model, and help designers improve their warning designs in Transportation CPS.

To our best knowledge, there are few mathematical models that predict human responses to speech warnings. Two major psycholinguistic models, the COHORT model and TRACE model, have described the mechanism of how human recognize and process spoken words in general. The COHORT model is a bottom-up verbal model that explains the lexical access for spoken word perception [18]. In the stage of activation, perception is influenced by auditory stimulation such that all words matching the perceived acoustic profile are activated, serving as a *cohort*. The selection stage refers to the process of selecting consistent input and eliminating candidate words that no longer match the input. Once the single candidate is isolated from the cohort, word recognition is accomplished. Unlike the COHORT model, the TRACE model is an interactive activated simulation model. The main feature of the model is the ability to describe the interaction of units including within-level inhibition and between-level facilitation [19]. The cascaded activation mode in the TRACE model enables the activation word-level processing units sooner after the activation of the feature-level processing units. The word with the most support from the bottom layers will increase its activation until only one candidate is left standing. These two cognitive models laid the significant foundation on understanding the mechanism of speech perception and processing.

However, the COHORT model and the TRACE model focus on the speech perception and recognition instead of human responses to speech. Therefore, they cannot be used to predict human performance in their responses to speech warnings. Meanwhile, both psycholinguistic models focus on general mechanism of speech processing rather than different characteristics of speech warnings so that they are not able to predict the effects of different characteristics of speech warnings on human responses. Moreover, neither COHORT nor TRACE model is a mathematical model. Mathematical models are indispensable to predict how human respond to speech warnings under the influence of warning characteristics in order to be applied in the design of transportation CPS. Therefore, new models are still needed to model how different characteristics of the speech warnings affect human responses with the mapping between the meaning of speech warnings and the target response actions.

The present work addresses this problem by developing a mathematical model to predict human responses to speech warnings in human–machine systems. This paper extended the model presented in [72] by integrating the algorithm of reinforcement learning in modeling the route choice in the processing of speech warnings and quantifying human reaction error rate and reaction time in speech warning responses. Three main speech-warning parameters are discussed: lead time, loudness, and signal word choice. As the causes of accident in reality can be very complex, the errors in initial responses and the slowed responses to warnings are two of major causes that leaded to traffic accidents. Therefore, accident rate is modeled as the outputs of the model with this two causes being considered and is tested with two empirical studies. In addition, the applications of the model were discussed in setting up the warning parameters to optimize the design of transportation cyber-physical system in terms of human performance. The interface of web-based software was proposed for designers as an easy-to-use technology to design different speech warning parameters associated with human performance.

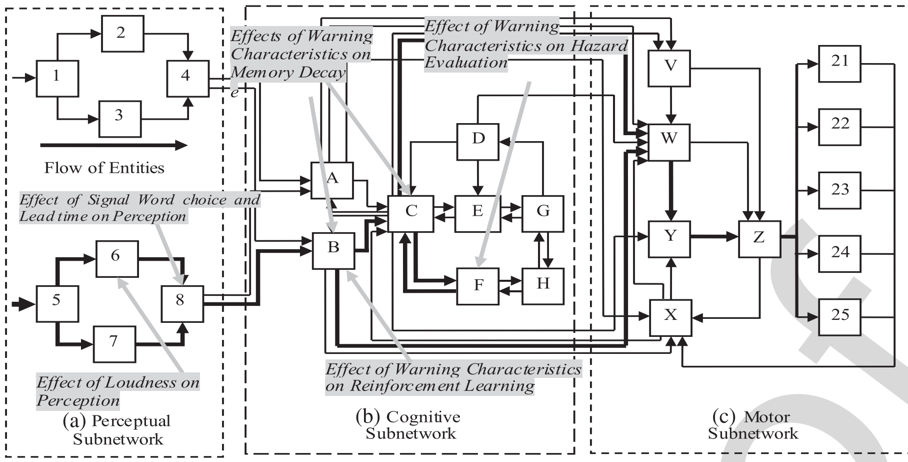
II. Modeling Mechanism and Model Enhancement

A. *Overview of Queuing Network-Model Human Processor (QN-MHP)*

Queuing Network-Model Human Processor (QN-MHP) is a computational architecture that integrates three discrete serial stages of human information processing (i.e., perceptual, cognitive, and motor processing) into three continuous subnetworks (see in Fig. 1). Each subnetwork is constructed of multiple servers and links among these servers. Each individual server is an abstraction of a brain area with specific functions, and links among servers represent neural pathways among functional brain areas. The neurological processing of stimuli is illustrated in the transformation of entities passing through routes in QN-MHP. Since this architecture was established, QN-MHP has been applied to quantify various aspects of human cognition and performance, such as human mental workload [20], and the reinforcement learning process [21]. In terms of the perceptual subnetwork, new equations have been integrated to model eye movements, and speed perception [22], [23]. The cognitive subnetwork has been improved to model textual information chunking [26], inhibition incompatible responses and choice reactions [24], dual task interference [25], and complex decision making [26]. Moreover, applications of QN-MHP indicate its success in modeling motor program retrieval [26], error corrections [25], bimanual coordination in typing tasks, and driver speed control [23], [26].

B. *Enhancements of Queuing Network-Model Human Processor (QN-MHP)*

In the present work, the mathematical model was proposed based on architecture of QN-MHP to predict human performance in speech warning responses with system operation tasks (e.g., driving a vehicle) based on neurological findings [34]–[38], [42]–[49]. Although several mathematical models based on the QN-MHP have been successfully built to predict driver behaviors such as speed and lateral control, the model to predict human responses to speech warning is still missing. The highlighted servers with labels in Fig. 1 illustrated the servers to be enhanced with the equations developed in the current work and the processing of speech warnings with the "Flow of Entities."



Perceptual Subnetwork	Cognitive Subnetwork	Motor Subnetwork
1. Common visual processing	A. Visuospatial sketchpad	V. Sensorimotor integration
2. Visual recognition	B. Phonological loop	W. Motor program retrieval
3. Visual location	C. Central executive	X. Feedback information collection
4. Visual recognition and location integration	D. Long-term procedural memory	Y. Motor program assembling and error detecting
5. Common auditory processing	E. Performance monitor	Z. Sending information to body parts
6. Auditory recognition	F. Complex cognitive function	21-25: Body parts: eye, mouth, left hand, right hand, foot
7. Auditory location	G. Goal initiation	
8.Auditory recognition and location integration	H. Long-term declarative & spatial memory	

Fig. 1. The enhanced servers of the QN-MHP architecture with the equations to be developed in the current work, and the general structure of QN-MHP (developed in [20], [21], [24], and [27]; and all of the published mathematical equations in QN-MHP can be found at: http://www.acsu.buffalo.edu/~seanwu/QNMHPMath/MathModelQNMHP_Online.htm).

In the speech warnings response task, the stimuli of speech warnings entered into the auditory perceptual subnetwork. The stimuli firstly arrived at Server 5, representing the middle and inner ear (common auditory processing). The parallel auditory pathways transmitted the auditory information through the neuron pathways from the dorsal/ventral cochlear nuclei to the inferior colliculus presented by Server 6 (auditory recognition) and from the ventral cochlear nucleus to the superior olivary complex represented by Server 7 (auditory location).

Then the auditory information was integrated at Server 8, representing the primary auditory cortex and the planum temporale (auditory recognition and location integration. The speech warnings with specific loudness and semantic features were then transmitted to the left-hemisphere posterior parietal cortex presented as Server B (phonological loop).

A route choice located at Server B with a shorter route directly connecting to Server W (motor programs retrieval) representing basal ganglia, and a longer route connecting to Server C (central executive) and Server F (complex cognitive function), and eventually leading to Server W. The shorter route represented a processing in emergent situations and the longer route involved the stage of hazard evaluation in less emergent situations. Those motor programs at Server W were then assembled at Server Y (motor program assembling and error detecting) and initialized at Server Z representing primary motor cortex, sending out the neural signals to body parts (Servers 21–25).

1) Modeling the Effect of Speech Warning Parameters on the Probability of Route Choice in Reinforcement Learning: The modeled routes in QN-MHP were presented in Fig. 1. As it showed at Server B, entities could choose one of the two routes to move to either Server C (long route) or Server W (short route). The division of the two routes was modeled with the route choice at Server B. Previous fMRI studies indicate two stages involved in processing warning signal words associated with hazards [28]. One stage is a rapid automatic activity and the other stage involves the activation of the hazard evaluation. The rapid automatic activity with a shorter response time to warnings could be represented by the shorter route (Route I) of warning responses learned through experiences in urgent situations [29], [30]. The other activity involving a hazard evaluation process could be represented by the longer route (Route II) of warning responses learned through experiences in non-urgent situations [35]. To process information with Route II, the human would take a longer time to respond as more servers were involved in this route. In the meantime, the human would have a lower error rate of responses since entities were processed through critical servers (Servers C and F) could correct errors to a certain degree.

The probability of choosing a route could be the result of learning from the connections of warning characteristics and associated hazards in daily life. Previous fMRI studies showed that people learned responses to auditory stimuli with a co-activation of the motor/premotor cortex and the primary auditory cortex [31]. As the neuron in motor and premotor cortex (Server W) fired repeatedly when the human processed associated warnings, the correlation of neuronal firing of connected cortical cells was translated into their connection strength [32]. At the beginning of the learning, entities of speech warnings with different loudness levels or signal words might have equal chances to enter either route. Then the probability of route choosing would be updated as humans learned from association between specific loudness levels/signal words and urgency of hazards.

Whether a situation was considered to be an emergency was determined by certain criteria of loudness levels and signal words. In terms of warning loudness, Blumenthal [33] reported that a 50% probability threshold of a startle response was 85 dB. Studies have shown the increasing of the acoustic stimuli intensity leads to an increase in response magnitude and amplitude, and a decrease in response onset latency [34]. For signal word choices, different signal words expressed different perceived urgency levels (Hollander & Wogalter, 2000). Therefore, speech warnings with its loudness higher than 85 dB or a particular signal word (e.g. "Danger") would represent an emergency situation.

Moreover, the incompatibility of warning loudness and word semantics indicating different hazard urgency levels took longer time for human to respond [28]. This incompatibility might result in entities traveling through a longer route (Route II) with higher chance in order to solve the incompatibility problem [35]. The probabilities of choosing route I (p_I) and route II (p_{II}) for speech warnings with certain loudness levels and signal words were obtained from the simulation results (see Q online learning algorithms in the Appendix).

2) Modeling the Effect of Speech Warning Characteristics on the Warning Perception, Memory Decay and Hazard Evaluation: The choices of servers and where to integrate equations were determined by the brain area are influenced by warning characteristics. Studies suggested loudness and signal word choice have significant effects on human behaviors [36]–[39]. It has been shown that the activation of lower auditory processing level increased with the sound level increased [40]. Therefore, the effect of loudness on speech warning perception was modeled at Server 6. The semantic features of signal words are recognized at the superior temporal sulcus, which was modeled at Server 8 [41].

Due to the interference caused by the speech warnings on the on-going tasks, memory decay may occur [42]. The effect of warning lead time on memory decay was modeled in the working memory system regarding auditory processing represented by Servers B and C. Previous fMRI studies indicated that hazard evaluation activated the medial prefrontal cortex, the inferior frontal gyrus, the cerebellum, and the amygdale [43], which were presented by Server F.

III. Mathematical Formulation of Modeling Mechanisms and the Enhancement of the QN-MHP

A. Modeling the Effects of Loudness and Signal Word Choice on Perceived Urgency and Annoyance of Speech Warnings

1) Modeling the Relationship Between Loudness and Perceived Urgency/Annoyance: The relations between changes in loudness and changes in perceived urgency can be quantified by the Stevens Power Law [38]. The loudness was reported having a positive relationship with urgency expression [44]. Therefore, the perceived urgency (U_L) and annoyance (A_L) as a function of warning loudness was modeled by the following equations:

$$\log(U_L) = m_U \log(L) + k_U + \varepsilon_1 \tag{1}$$

$$\log(A_L) = m_A \log(L) + k_A + \varepsilon_2 \tag{2}$$

where L denotes the loudness level and m and k quantify the relationship between perceived value and objective loudness change. The relationship between intensity and perceived urgency/annoyance was quantified [44]. The Stevens' power law states that the loudness (L) is proportional to $I^{0.3}$, where I is the sound intensity [45]. Therefore, the parameters are quantified as: $m_U = 1.33$, $m_A = 1.45$, $k_U = -0.64$, $k_A = -0.91$. ε_1 and ε_2 are normally distributed random factors following distribution [0, 0.7] and [0, 0.86], respectively [36].

2) The Relationship Between Signal Word Choice and Perceived Urgency: Considerable research efforts have been indicating a stable relationship between signal word choice and perceived urgency. Hollander & Wogalter (2000) reported ratings in carefulness expressed in a descending order by the following five signal words: deadly, danger, warning, caution and notice. Other studies have found similar results. These words covered a wide range of urgency ratings and have been studied before in detail (Barzegar & Wogalter, 1998; Hollander & Wogalter, 2000) using the word "notice" rather than "note." The perceived urgency of "danger," "caution," and "notice" spoken by a female voice on a 100 points scale are quantified as 90.53,72.40, and 46.81 [44].

B. Modeling the Error Rate in Speech Warning Responses

Speech warning parameters have different influence on speech warning response error rate in different stages of speech warning responses. When humans processed speech warnings through route I, the error rate was mainly influenced by the effects of loudness and signal words on speech warning perception. When speech warnings were processed through route II, the error rate in the speech warning responses was also influenced by the effects of lead time on potential memory decay of the speech warnings and hazard evaluation.

1) Modeling the Effect of Loudness and Signal Word Choice on Error Rate: Errors in speech warning responses could result from the shortcoming of perception, memory, cognition and the failure in motor execution [46]. Errors in speech warning responses include no responses to correct warnings (e.g., failures in recognizing speech warnings and misjudging hazards associated with warnings) or incorrect responses to warnings (e.g., accelerating instead of braking towards a forward collision). The error rate (I_E) is modeled as a function of the speech warning loudness and signal word choice and the corresponding probability of route choices. A warning with higher urgency is correlated with higher arousal strength, which may result in a startle reflex and lead to a higher chance of poorly processing the warning signal words [28]. This autonomic activity can be represented as entities traveling through route I with a higher chance of making errors such that entities are not processed in critical Server C and Server F. Both loudness and semantic features relevant to the expressed urgency of the speech warnings have influence on error rate in the perception of speech warnings [47]. Also, a positive correlation between loudness and error rate was found in an empirical study [48]. The error rate in route I is then modeled with a positive correlation with perceived urgency expressed by word loudness and word semantics.

The entity processed through route II involves the central executive and hazard evaluations at Servers C and F. The effect of loudness on error in response would decrease after the entity passed the phonological loop due to the decay of the echoic memory [52]. Further processing of the entity led to pattern recognition or semantic analysis of the speech warnings (at Server C) and the corresponding hazard was evaluated in the decision making stage (at Server F) [28], [49]. Therefore, the error rate in route II was modeled with a correction of errors brought in by the loudness and semantic properties of the speech warnings in the perceptive stage of speech warnings.

In summary, the error rate $(I_{E,i})$ of route i $(i = I$ or $II)$ is modeled with the following equation (3) with the perceived urgency (U_L) and annoyance (A_L) of speech warnings due to different loudness levels, and the perceived urgency of speech warnings due to different signal words (U_S). Since there is no difference of perceived annoyance due to different signal words (A_S), it is not inputted in modeling the error rate

$$I_{E,i} = \begin{cases} (U_L + U_S) \times 0.5, & i = I \\ (U_L - A_L) \times 0.5, & i = II \end{cases} \tag{3}$$

where L is the speech warnings loudness and S is the signal words. U_L and A_L are the perceived urgency and annoyance of warning loudness obtained from (1) and (2); U_S is the urgency of signal word choice. According to the perceived urgency for signal word scales, the perceived urgency for word semantics (U_S) is 0.90, 0.72 or 0.47 for signal words "Danger," "Caution," "Notice," respectively [44].

The overall error rate in the responses to speech warnings is then modeled by adding up the error rate with the probability in each route. The effect of speech warning parameters on route choice error rate (I_E) can be modeled as the combined effect of the speech warning loudness and signal word choice:

$$I_E = \sum_{i=1}^{2} I_{E,i} \times p_i \tag{4}$$

where $I_{E,i}$ denotes the error rate when a speech warning travels through route i. p_i denotes the probability of information processing through route i.

Then the equation (4) for the effect of speech warning loudness and signal word choice on error rate (I_E) is updated by the following general equation:

$$\begin{aligned} I_E = {} & (L^{m_U} \times 10^{k_U - 2} + U_S) \times 0.5 \times p_I \\ & + \left((L^{m_U} \times 10^{k_U-2} - L^{m_A} \times 10^{k_A - 2}\right) \times 0.5 \times p_{II} \end{aligned} \tag{5}$$

where L denotes the loudness level in dB. U_S is the perceived urgency level with different signal word choice. p_I and p_{II} are probabilities of choosing route I (the shorter route) and route II (the longer route) respectively obtained from the simulation results of the reinforcement learning in Appendix. m_U and k_U are parameters to quantify the power law of perceived urgency and loudness. m_A and k_A are parameters to quantify the power law of perceived annoyance and loudness.

2) Modeling the Impact of Lead Time on Error Rate: Drivers tend to respond to the speech warning when the corresponding hazard is within sight [13]. When there is a relatively long lead time before the actual hazard occurrence, the human may perform normal operations and monitor the situation. Therefore, the memory of the speech warnings may decay and the corresponding accuracy rate of upcoming hazard estimation may increase the error rate in responses to speech warnings.

The probability of information retrieving (p) is modeled as a function of time (t) starting from the information presented to humans in [42] as follows:

$$p = e^{\text{at}}, [42] \tag{6}$$

where $a = -0.02$ based on parameter settings of MHP [50].

In the proposed speech warning responses model, the effect of lead time on memory decay (I_{MD}) is computed at Servers B and C in QN-MHP, representing the working memory system regarding auditory information processing

$$I_{MD} = \frac{1}{e^{-0.02t}\text{lead}}. \tag{7}$$

In the above equation, t_{lead} denotes the lead time for speech warning responses.

In terms of hazard estimation, a human will react to speech warnings when a perceived hazard reaches a certain threshold. The effect of hazard evaluation accuracy on error rate (I_H) can be modeled by the difference between the perceived value and the actual value of the hazard in the following equation:

$$I_H = \frac{H_p}{H_0} \tag{8}$$

where H_p denotes the perceived value of hazard and H_0 denotes the actual value of hazard.

In summary, the error rate (r) in speech warning responses is extended by adding the effects of loudness and signal word choice modeled in (5), and the effect of lead time modeled in (7) and (8) as follows:

$$r = I_E + I_{MD} \times I_H + \varepsilon_3 \tag{9}$$

where I_E denotes the error from signal word perception and recognition under the effect of speech warning loudness and signal word choice, I_{MD} denotes the error from memory decay, I_H denotes the error from hazard location estimation. ε_3 is a random factor following normal distribution [0, 0.1] [51].

C. Modeling the Reaction Time in Speech Warning Responses

The reaction time was defined as the time duration from the time the speech warning occurs to the time the human starts to react. As assumed in QN-MHP, entity processing time at an individual server is independent of arrivals of entities, and routing is independent of the state of the system. Therefore, the reaction time of a speech auditory stimulus can be modeled by summarizing the processing time of all the servers on the route.

Consequently, the reaction time (RT_i) to speech warnings through route i is modeled as:

$$RT_i = \begin{cases} T_5 + T_6 + T_8 + T_B + T_W + T_Y + T_Z, & i = I \\ T_5 + T_6 + T_8 + T_B + T_C + T_F + T_C & \\ \quad + T_W + T_Y + T_Z & i = II \end{cases} \tag{10}$$

where T_k is the processing time of auditory stimulus at Server k. The processing time of servers in perceptual, cognitive, and motor subnetwork are 42 ms, 24 ms, and 18 ms [24].

The effect of loudness on reaction time is modeled in the initial processing of auditory stimuli in Server 6

$$T_6 = \frac{T_{6(0)}}{U_L} \tag{11}$$

where $T_{6(0)}$ is the initial entity processing time in Server 6 and U_L denotes the effect of loudness on perceived urgency.

The effect of signal word choice on reaction time can be modeled by the following equation:

$$T_8 = \frac{T_{8(0)} \times n_i}{U_s} \tag{12}$$

where $T_{8(0)}$ is the entity processing time in Server 8 and n_i is the number of words in the ith speech warning. U_s denotes the urgency level expressed by the initial words (e.g., signal words) in the speech warnings.

All in all, the equation (10) for modeling reaction time of speech warnings through route i is updated as:

$$\begin{aligned} RT = &\left(T_5 + \frac{T_{6(0)}}{U_L} + \frac{T_{8(0)}}{U_s} + T_B + T_W + T_Y + T_Z\right) \\ &\times p_I + \left(T_5 + \frac{T_{6(0)}}{U_L} + \frac{T_{8(0)}}{U_s} + T_B + T_C + T_F \right. \\ &\left. + T_C + T_W + T_Y + T_Z\right) \times p_{II} + \varepsilon_4. \end{aligned} \tag{13}$$

In the above, T_k denotes the processing time of the auditory stimulus at Server k $(k = 5-8, B, C, F, W-Z)$. U_L is the perceived urgency level with different levels of loudness. p_I and p_{II} are probabilities of choosing route I (the shorter route) and route II (the longer route), respectively. ε_4 is a normally distributed random factor following distribution [0, 0.3] [13].

D. The Application of Speech Warning Response Model in Driving and Warning Responses

The following section presents the application of the speech warning responses model in modeling human responses to speech warnings in Transportation CPS systems (e.g., in-vehicle information systems and connected vehicle communication systems). Warning responses in a driving task include the releasing of the accelerator pedal when drivers are accelerating and the change in braking pedal when drivers are already braking (i.e., foot on brake pedal) or on their way to brake (i.e., releasing the accelerator). The parameters of speech warnings are loudness and signal word choice, as well as lead time. The drivers tend to respond once the speech warning begins when they hear the signal words (e.g. "Notice," "Caution," "Warning," and "Danger"). QN-MHP was used to estimate the reaction to the speech warnings starting from perceiving the information from speech warnings to transmit neural signals to the foot server (Server 25).

1) The Hazard Evaluation in the Driving and Speech Warning Responses Tasks: When the speech warnings are presented to a driver, he/she will continuously evaluate the potential hazard based on the information obtained from visual perception and from speech warnings (e.g., estimated distance). Previous work studied the effects of motion factors (e.g., optical flow rate, optical density of texture and edge rage) and cognitive factors (e.g., perceived time, actual speed) on the traversed distance estimation [52]–[54]. Traveling speed had a significant effect on distance estimation, with slower speed resulting in more accurate distance estimation. The relationship between actual distance and estimated traversed distance (D_P) was quantified with Steven's power law [55]

$$D_P = D_0^{b^v} \tag{14}$$

where D_0 denotes the actual distance between the current position of warning receiving vehicle and the potential hazard location when speech warning is presented, while v denotes the instant speed $(b = 0.955)$ [55]. Based on the definition, the actual distance D_0 is modeled as:

$$D_0 = v_0 t_{\text{lead}} + \frac{1}{2} a_0 t_{\text{lead}}^2. \tag{15}$$

When the perceived distance is shorter than the minimum safety headway, drivers may react to the speech warnings directly. Otherwise, drivers continue to drive and react to speech warnings until perceived distance (D_p) reaches the threshold $(D_p = D_h)$. The hazard evaluation effect on crash rate is modeled as

$$I_H = \frac{D_h}{D_0} = D_0^{b^{v(t)}-1}. \tag{16}$$

The instant speed (v) and acceleration (a_t) at time t is modeled in [23] as follows:

$$v(t) = v_0 + a_t(\Delta t), [23] \tag{17}$$

where v_0 denotes the initial speed and a_t denotes the acceleration at time t.

The constant rate of deceleration $(a_t(\Delta t))$ is modeled in [56] as follows:

$$a_t(\Delta t) = \frac{k}{2} \times \phi \times \frac{\dot{\theta}}{\theta}, [56] \tag{18}$$

where ϕ is the global optic flow rate of the textured ground surface, a proportion of speed as long as eye height is constant. The global optic flow rate is constant in a braking task. The ratio $\dot{\theta}/\theta$, where θ and $\dot{\theta}$ are the optical angle and rate of expansion of approached object, respectively, is approximately equal to v/S. Therefore, the ideal deceleration can be expressed in terms of the optical variable by substituting ϕ for v and $\dot{\theta}/\theta$ for

v/S. Novices tended to initiate emergency braking earlier than necessary when initial speed was slow and to a lesser extent, which brought in a parameter k of driving experiences ($0 < k < 1$). The parameter k is quantified by the annual mileage divided by a maximum value of annual mileage in general.

The ratio of the object's optical angle to rate of expansion of approached object ($\dot{\theta}/\theta$) specifies the time-to-collision (TTC) with the object as long as the current velocity is held constant. The ratio is modeled in [57] as follows:

$$\frac{\dot{\theta}}{\theta} = \mathrm{TTC}, [57]. \tag{19}$$

The perceived time-to-collision (TTC_p) will be affected by the existence of the lead vehicle. TTC is the actual time to collision that the vehicle will be able to avoid a collision without exceeding the assumed maximum deceleration, which is represented as t_{lead} as above

$$\frac{\dot{\theta}}{\theta} = \mathrm{TTC}_p = t_{\mathrm{lead}} \times \exp(LV). \tag{20}$$

In the above, LV is a dichotomous variable of the lead vehicle in order to model the effect of the lead vehicle on TTC_p ($0 =$ without lead vehicle; $1 =$ with lead vehicle).

In summary, the effect of hazard evaluation on crash rate is modeled as:

$$\begin{aligned} I_H &= \frac{D_p}{D_0} = D_0^{b^{v_0 + \frac{k}{2} \times \phi \times \mathrm{TTC}_p} - 1} \\ &= \left(v_0 t_{\mathrm{lead}} + \frac{1}{2} a_0 t_{\mathrm{lead}}^2 \right)^{b^{v_0 + \frac{k}{2} \times \phi \times t_{\mathrm{lead}} \times \exp(LV)} - 1}. \end{aligned} \tag{21}$$

2) Modeling the Crash Rate in Speech Warning Responses: The modeling of crash rate has to consider the additional impact of warning lead time. Even if the driver makes correct responses, lack of time to respond will also result in accidents. When the lead time is shorter than the minimum brake-to-maximum response time ($t_{\min}$), the drivers may not avoid the collision even when they correctly respond immediately. Therefore an effect of lead time on crash rate is modeled as:

$$t_{\min} = \frac{v_0}{a_{\mathrm{average}}} + RT = \frac{v_0}{\frac{1}{2}|a_0 + a_{\max}|} + RT \tag{22}$$

$$I_{LT} = \frac{t_{\min}}{t_{\mathrm{lead}}}. \tag{23}$$

The impact of parameters (i.e., loudness and signal word choice) of speech warning on crash rate (R_{crash}) can be modeled by combining Equations (5), (21), (23) as follows:

$$R_{\mathrm{crash}} = I_E + I_{MD} \times I_H \times I_{LT} + \varepsilon_5 \tag{24}$$

where I_E denotes the error from signal word perception and recognition under the effect of speech warning loudness and signal word choice, I_{MD} denotes the error from memory decay, I_H denotes the error from hazard location estimation. I_{LT} denotes the effect of lead time on crash rate. ε_5 is a normally distributed random factor following normal distribution [0, 0.05] [14].

IV. The Validation of the Speech Warning Response Model

In order to validate the speech warning responses model, the following section provides the prediction results of two experimental studies in terms of driver responses to speech warnings. The first study conducted by our research group studied the effect of lead time on driver responses to speech warnings. In order to validate the model, the model predictions for response time and crash rate are shown and compared to experimental data. The second study from a published work examined the effect of loudness and signal word choice of warnings on rear-end collision [58]. Due to a lack of detailed information in the second study, the lead time and hazard evaluation was assumed to have no additional effect on modeling crash rate. The model predictions for crash rate and subjective ratings for perceived urgency and annoyance are shown and compared to experimental data. To validate the speech warning response model, the comparability of model predictions and experimental results were quantified by the Pearson correlation coefficient (R squared) as well as the root mean-squared error (RMSE).

A. Experiment 1

The first experiment involving a driving simulator was conducted to study the impact of lead time on human responses to speech warnings.

1) Participants: Thirty-two participants (24 males, 8 females),draftrules with ages ranging from 18 to 26 years participated in the study. All of them were licensed drivers and had normal or corrected-to-normal vision. None of the drivers had previously participated in any simulator or crash avoidance studies.

2) Apparatus: A STISIM driving simulator (STISIMDRIVE M100K, Systems Technology Inc, Hawthorne, CA) was used in the study. It comprises a Logitech Momo steering wheel with force feedback (Logitech Inc, Fremont, CA), a throttle pedal, and a brake pedal. The STISIM simulator was installed on a Dell Workstation with a 256 MB PCIe $\times$ 16 nVidia graphics card, Sound Blaster X-Fi system, and Dell A225 Stereo System. Driving scenarios were presented on a 27-inch LCD with 1920×1200 pixel resolution. A speaker in front of the participant provided auditory messages in a digitized human female voice with a speech rate of ~150 words/min and loudness level of ~70 dB. Another speaker provided driving sound effects with a loudness level of ~55 dB.

The behavioral measures (time elapsed (s), speed (m/s), acceleration (m/s^2), and distance to the initial location where the scenario starts (m) were automatically collected from the driving simulator and outputted to another identical Dell Workstation. This computer calculates the time to collision (TTC) in real time based on the vehicle's speed and acceleration. When the calculated TTC reached the designed value, the warning would be issued.

3) Scenario Setting: The speech warning would sound before the appearance of the hazard. Each speech warning started with a signal word "Caution" and followed by a description of the collision scenario presented (e.g., A vehicle at your front-left is running red light). The collision scenario description

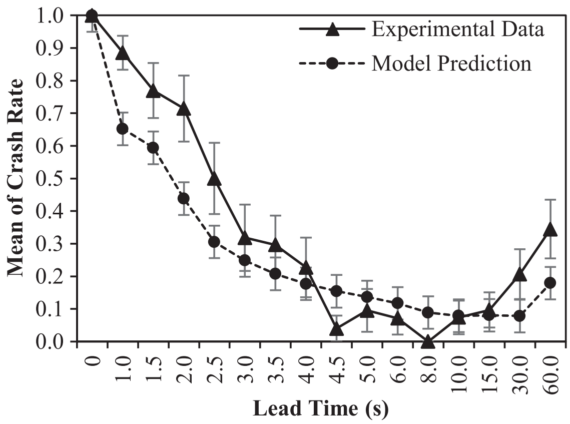


Fig. 2. The comparison of model prediction of crash rate with experiment 1 data (Error bars: +/−1 SE).

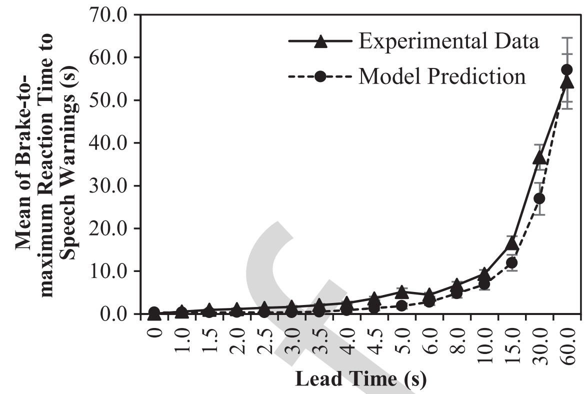


Fig. 3. The prediction of brake-to-maximum response time to the speech warnings (Error bars: +/−1 SE).

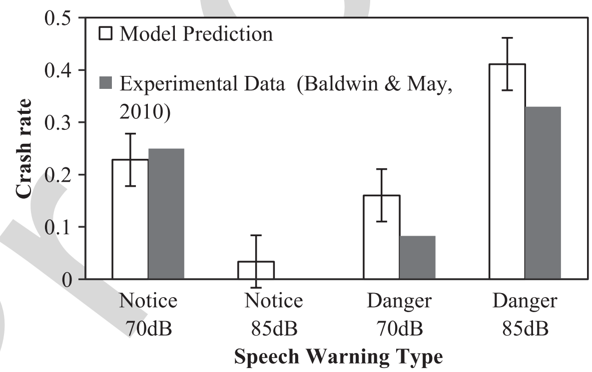


Fig. 4. The comparison of model prediction of crash rate with experiment data (no data of standard error reported in the experiment).

comprised the hazard location and event, which provided the driver with specific information to eliminate any confusion.

The test block used a two-lane (in each direction) urban environment with traffic lights and road signs. There were running vehicles moving in each direction. Speed limit signs with a constant speed limit of 45 mph (20 m/s) were displayed 200 feet (61 m) in front of the driver. Participants were instructed to adjust their speed within the range from 40 mph (18 m/s) to 50 mph (22 m/s) as if they were driving a vehicle in the real world. Sixteen collision scenarios were designed and programmed. A lead vehicle would run at the same speed as the subject vehicle. In order to investigate drivers' responses to speech warnings, their sights of the collision scenario were blocked by other vehicles, and participants could only rely on the warnings to learn about the upcoming collision events.

4) Experiment Design: The current experiment adopted a one-factor experiment design with lead time as an independent variable and collision rate and brake-to-maximum response time as dependent variables. The lead time had 16 levels (0 s, 1 s, 1.5 s, 2 s, 2.5 s, 3 s, 3.5 s, 4 s, 4.5 s, 5 s, 6 s, 8 s, 10 s, 15 s, 30 s, and 60 s). When the lead time was 0, the warning sounded at the same time when the collision event happened. Each subject would go through sixteen collision events with sixteen levels of lead time assigned to each event. The orders of levels of lead time and events were randomized. The normal messages were randomly assigned during the experiment, as long as they did not cause interference with the broadcasting of speech warnings.

The first dependent variable was collision, which specified whether there was collision between a subject's vehicle and a hazard vehicle. The collision rate was then calculated as the percentage of collisions for each level of lead time. Brake-to-maximum response time represented the time period from the present of warnings until drivers reaching the maximum deceleration in the braking responses.

5) Results: The model prediction for crash rate with speech warnings of different lead time levels is shown in Fig. 2. The RMSE was 0.13 with an R square of 0.94. For the brake-to-maximum response time to the speech warnings, Fig. 3 showed the model prediction comparing the experimental results had an R square of 0.97 and RMSE of 3.17.

B. *Experiment 2 (Baldwin & May, [64])*

The second experimental study examined the effect of loudness and signal word choice of in-vehicle collision warnings on driver responses [64]. Thirty participants were recruited to drive through five different scenarios containing five different hazard events. Speech warnings consisted of the signal word "Notice" or "Danger" presented at either 70 or 85 dBA. The driving sound effects were presented with a loudness level of 55 dB. The crash rate with different warnings and subjective rating of perceived urgency and annoyance were reported.

Due to a lack of detailed information regarding collision event scenario and driver responses, the lead time was set up to be long enough for effective responses in this study since there is no lead time reported ($I_{LT} = 1$). The model prediction for crash rate with different speech warnings is shown in Fig. 4. The RMSE was 0.06 with an R square of 0.90. Fig. 5 shows the model prediction of rating of urgency and annoyance for signal word. The R square of perceived urgency prediction is 1.00 with RMSE of 1.49. The R square of annoyance is not calculated since there is no differences among annoyance ratings of signal words [42].

V. The Application of Prediction of Human Performance in Developing Speech Warnings

Speech warnings in Transportation Cyber-Physical Systems are designed to improve driver safety by providing information about upcoming hazards in an appropriate way so as to

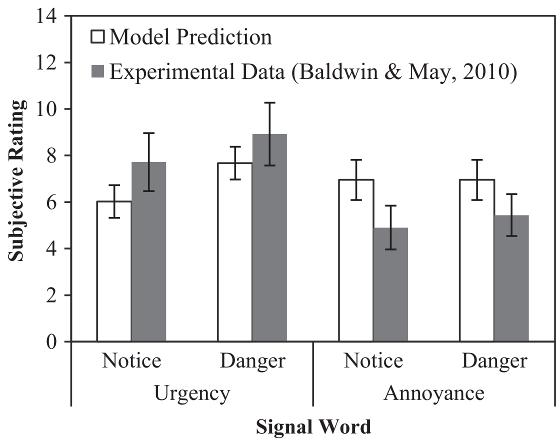


Fig. 5. The prediction of rating of urgency and annoyance for signal words (Error bar: standard deviation).

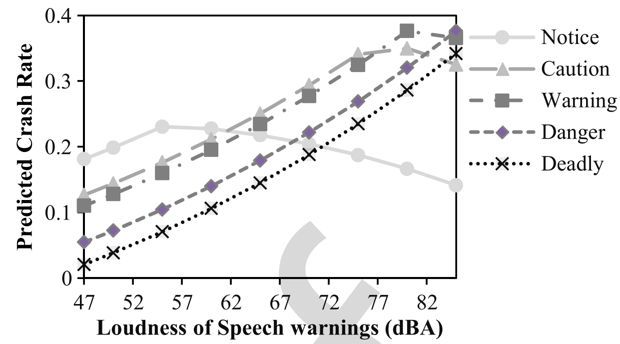


Fig. 6. Predicted crash rate with speech warnings presented at different loudness levels with different signal words.

give drivers enough time to respond. Previous work mainly studied speech warning characteristics through experimental approaches [19], [64]. The developed model in the current work makes it easier for designers to obtain the effects of different speech warning parameters associated with human performance. In particular, the warning lead time, loudness and signal word choice can be optimized by applying the developed model to simulate human performance. Taking the intelligent transportation system as an example, the crash rate will serve as the objective index of potential safety benefit of the speech warnings.

Based on abovementioned modeling results, the model predicted that crash rate would vary with different combinations of lead time, loudness and signal words. Equation (21) is applied to quantify collision rate under different combinations of loudness and signal words with the common noise loudness level of 55 dBA. The threshold of intelligibility (TI) is at the loudness level of 47 dBA, which was defined as the "level at which the listener is just able to obtain without perceptible effort the meaning of almost every sentence and phrase of the connected discourse" [65]. A human, therefore, will not fully recognize and understand warnings with loudness levels below this threshold. The predicted impact of loudness and signal words on crash rate shown in Fig. 6 illustrate the loudness level with range from 47 to 85 dBA with a lead time of zero as an example. The best loudness level to present the signal word "Notice" is 85 dBA, whereas the best loudness level for other signal words is 47 dBA. It is implied that the combination of speech warnings with an intermediate urgency level brought the most safety benefits.

The joint effect of lead time and warning loudness level is shown with the signal word "Caution" (see Fig. 7) as an example. Likewise, the joint effect of lead time and warning signal words is shown with the loudness level of 70 dBA (see Fig. 8). The predicted crash rate has a descending trend as a function of lead time regardless of the impact of loudness level and signal words. Generally speaking, it suggested that early warnings resulted in lower crash rates than did late warnings. As it is shown in Fig. 8, an abrupt decrease of collision rate appeared with longer lead time when the warning was relatively late; the rate of such decrease tended to slow down when the warning was relatively early. The differences in crash rate between different loudness levels and signal words reduced when the lead time was longer. In other words, the impact of loudness and the signal word choice on human responses will decay with the processing of the speech warnings.

Future software can be designed based on the developed models in this work to specify the loudness and signal word choice of speech warnings in the Transportation Cycle-Physical Systems. A sample interface is shown in Fig. 9. With the loudness, signal words and number of words in the speech warnings inputting into the software, the designers of the warning system will be able to obtain the objective parameters regarding human responses, including the predicted crash rate and brake-to-maximum warning response time. Moreover, the subjective rating of the speech warnings could also be obtained by applying this model.

VI. Discussion

In this modeling work, mathematical equations were built within the framework of the Queuing Network Model Human Processor (QN-MHP) to predict human performance in speech warning responses, including human error rate and response time with different warning characteristics. No free parameters were used in the parameter setting. The validation of the model with two laboratory studies indicated its relatively good ability to predict performances in speech warning response with high correlations with behavioral data from two experiments [64].

This work is one of a few mathematical models with analytic solutions in the field of human speech processing. Previous modeling work has explored theories that account for the experimental data of word recognition and speech comprehension [66, 67]. In the review of word recognition models, most modeling work focuses on the mechanism of speech recognition with either verbal models (e.g., COHORT) or simulation models with descriptions of theory implemented in computer programs (e.g., TRACE) [22], [68, 69]. Compared to verbal or simulation models, the conciseness and rigor of mathematical models allows an easier implementation for different systems regardless of the computer language used in the system.

More importantly, few computational models focused on the prediction of human performance in speech warning responses

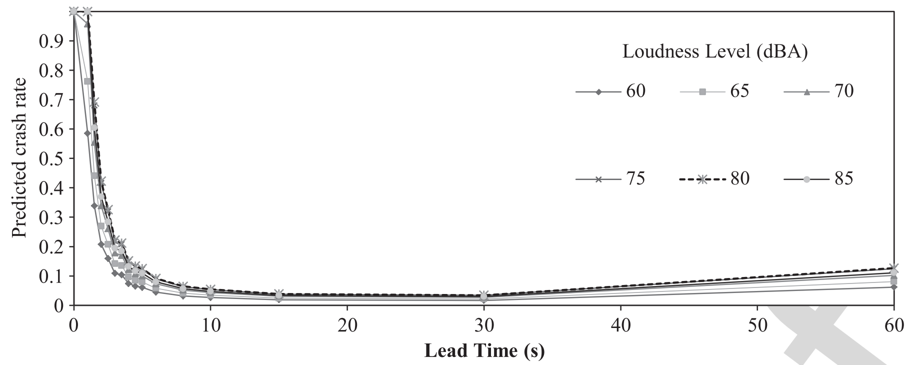


Fig. 7. Predicted crash rate with speech warnings presented at different lead time level and loudness levels (using signal words "Caution").

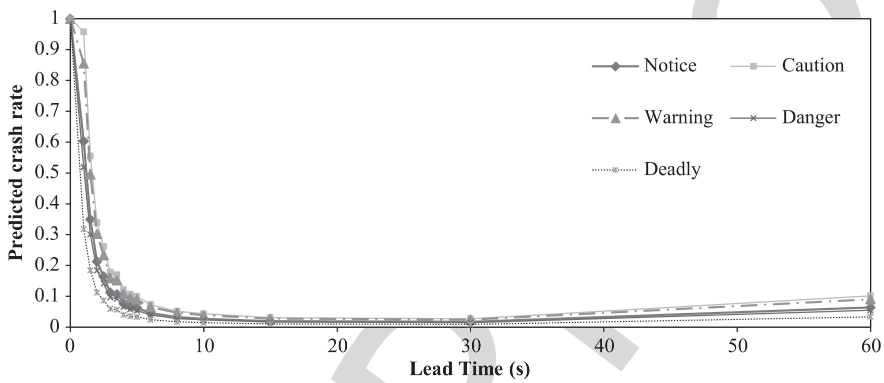


Fig. 8. Predicted crash rate with speech warnings presented at different lead time level and different signal words (at loudness levels = 70 dBA).

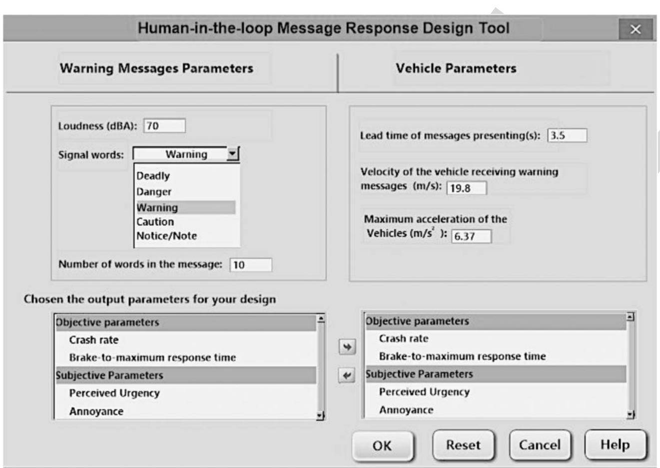


Fig. 9. The sample interface of the software with the application of the model.

and considered the characteristics of speech warnings. The neighborhood activation model focused on the prediction of the unique time point of word recognition [70]. The models that do model human performance (e.g., [71]) haven't predicted response error rate. In the current work, humans were responding to warnings associated with driving tasks rather than that to isolated words. In this case, the modeled process involves the hazard evaluation associated speech warnings and the selection of proper manual responses with the effects of warning characteristics being modeled. This different emphasis on human response modeling is important in the design of transportation CPS, since such systems have to consider how human respond to speech warnings by changing their operating behavior under the influence of different warning characteristics.

Although this study was carefully prepared, there are still several limitations. First of all, the model was mainly validated with accident rates and response time since the published work only reported the accident rate as the objective index of warning response performance. Further work is needed to validate the detailed levels of the proposed model. Secondly, although the proposed mathematical model provides a promising tool to predict the effects of loudness of speech warnings on human performance, the influence of other acoustic properties, like frequency and pitch, and the threshold of intelligibility were not modeled. For example, the warning presented with a higher pitch (e.g., female voice) may have a different impact on human performance and subjective rating on warning urgency than that of a lower pitch (e.g., male voice). Meanwhile, there might be interactions between signal word choice and other acoustic factors. The current work assumed that the perceived urgency expressed by different signal words is relatively stable, but the perceived urgency might vary with the signal words presented at different pitch and frequency levels. To enhance the model in predicting speech warning acoustics and semantic properties on human behaviors, further work is needed to model the interaction among acoustic properties and the interaction between signal word choice and other acoustic properties. Furthermore, the current QN-MHP model did not account for individual differences, but may significantly contribute to the model application. For example, although the model predicted optimal lead time, loudness level and signal word for speech

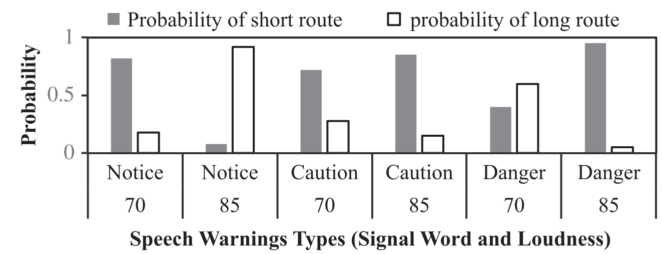


Fig. 10. The simulation results of route choice for warnings with different loudness level and signal words.

warnings, personality (e.g., aggressive vs. conservative drivers) may affect a driver's responses to speech warnings. Ideally, future model should consider individual differences and provide different system design suggestions according to individual characteristics instead of an average driver.

APPENDIX
THE Q ONLINE LEARNING ALGORITHM AND MODELING OF LEARNING PROCESS

The Q online learning algorithm will be integrated with the QN-MHP to model the learning in route choice under the influence of warning loudness and word choice. The effect of speech warning parameters on reaction time ($I_{RT,i}$) and response error rate ($I_{E,i}$) is then modeled with the different route choices in the information processing. As it presented in the following equations (Equations 9.2 and 9.3,) [21], the choice of route is based on the updated Q value $Q^{t+1}_{(i,j)}$ in each transition:

$$Q^{t+1}_{T(i,j)} = Q^{t}_{T(i,j)} + \varepsilon\left\{r'_t + \gamma \max_k\left[Q^{t}_{T(i,k)}\right] - Q^{t}_{T(i,j)}\right\} [21] \tag{25}$$

$$Q^{t+1}_{E(i,j)} = Q^{t}_{E(i,j)} + \varepsilon\left\{r''_t + \gamma \max_k\left[Q^{t}_{E(i,k)}\right] - Q^{t}_{E(i,j)}\right\} [21] \tag{26}$$

where $Q^{t+1}_{(i,j)}$ is the online Q value if entity routes from server i to server j in $t+1$th transition. $\max_k[Q^{t}_{(i,k)}]$ denotes the maximum Q value routing from server j to next k servers ($k \le 1$); r_t is the reward; γ is the discount parameter of routing to next server ($0 < \gamma < 1$). The time-saving reward (r'_t) is modeled as $r'_t = (1/w_q) + \mu_{j,t}$, where w_q is the waiting time in the queuing at the server; the error-saving reward r''_t is modeled as $r''_t = (1/(N_{\mathrm{error}(j,t)} + 1))$, where $N_{\mathrm{error}(j,t)}$ is the number of action errors of the previous entities made in the next server j at tth transition

$$N_{\mathrm{error}(j,t)} = N_{\mathrm{error}(j,t)} + 1 \times L/100 \times U_S.$$

Both $Q^{t+1}_{E(i,j)}$ and $Q^{t+1}_{T(i,j)}$ will contribute to the survival chance when human respond to warnings toward a potential hazard. Therefore, the choice of routes is determined by the sum of two Q values. Currently, it is assumed that Q value of the error-saving reward and the Q value of the time-saving reward has the same priority. If $Q^{t+1}_{E(i,j)} + Q^{t+1}_{T(i,j)} > Q^{t+1}_{E(i,k)} + Q^{t+1}_{T(i,k)}$, the entity will choose server j; if $Q^{t+1}_{E(i,j)} + Q^{t+1}_{T(i,j)} < Q^{t+1}_{E(i,k)} + Q^{t+1}_{T(i,k)}$, the entity will choose server k; and if $Q^{t+1}_{E(i,j)} + Q^{t+1}_{T(i,j)} = Q^{t+1}_{E(i,k)} + Q^{t+1}_{T(i,k)}$, the entity will choose the next server (j or k) randomly. The simulation results of probability of route choices is shown in Fig. 10.

REFERENCES

[1] NHTSA, "2011 Traffic Safety Facts Data Summary Booklet," U.S. Dept. Transp., Washington, DC, USA, 2013. [Online]. Available: http://www-nrd.nhtsa.dot.gov/Pubs/811755DS.pdf.

[2] D.-Y. Cheng *et al.*, "The optimal sampling period of a fingerprint positioning algorithm for vehicle speed estimation," *Math. Prob. Eng.*, vol. 2013, 2013, Art. no. 306783.

[3] J. Santa, A. F. Gómez-Skarmeta, and M. Sánchez-Artigas, "Architecture and evaluation of a unified V2V and V2I communication system based on cellular networks," *Comput. Commun.*, vol. 31, no. 12, pp. 2850–2861, Jul. 2008.

[4] J. Edworthy and R. Hards, "Learning auditory warnings: The effects of sound type, verbal labelling and imagery on the identification of alarm sounds," *Int. J. Ind. Ergonom.*, vol. 24, no. 6, pp. 603–618, Oct. 1999.

[5] K. Momtahan, R. Hetu, and B. Tansley, "Audibility and identification of auditory alarms in the operating room and intensive care unit," *Ergonomics*, vol. 36, no. 10, pp. 1159–1176, Oct. 1993.

[6] R. D. Patterson and C. A. Authority, *Guidelines for Auditory Warning Systems on Civil Aircraft*. London, U.K.: Civil Aviation Authority, 1982.

[7] J. Serrano, L. L. Di Stasi, A. Megias, and A. Catena, "Effect of directional speech warnings on road hazard detection," *Traffic Injury Prev.*, vol. 12, no. 6, pp. 630–635, Dec. 2011.

[8] K. R. Laughery, "Safety communications: Warnings," *Appl. Ergonom.*, vol. 37, no. 4, pp. 467–478, Jul. 2006.

[9] R. G. Mortimer, "Effectiveness of a warning as measured by behavior change," *Proc. Hum. Factors Ergonom. Soc. Annu. Meet.*, vol. 51, no. 9, pp. 574–577, Oct. 2007.

[10] M. S. Wogalter, M. J. Kalsher, and B. M. Racicot, "Behavioral compliance with warnings: Effects of voice, context, and location," *Safety Sci.*, vol. 16, no. 5/6, pp. 637–654, Aug. 1993.

[11] M. S. Wogalter, M. J. Kalsher, and R. Rashid, "Effect of signal word and source attribution on judgments of warning credibility and compliance likelihood," *Int. J. Ind. Ergonom.*, vol. 24, no. 2, pp. 185–192, May 1999.

[12] J. Wan, C. Wu, and Y. Zhang, "The effects of lead time of collision warning messages on driver performance," *Proc. Hum. Factors Ergonom. Soc. Annu. Meet.*, vol. 58, no. 1, pp. 2088–2092, Sep. 2014.

[13] G. Abe and J. Richardson, "The effect of alarm timing on driver behaviour: An investigation of differences in driver trust and response to alarms according to alarm timing," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 7, no. 4/5, pp. 307–322, Jul.–Sep. 2004.

[14] J. D. Lee, D. V. McGehee, T. L. Brown, and M. L. Reyes, "Collision warning timing, driver distraction, and driver response to imminent rear-end collisions in a high-fidelity driving simulator," *Hum. Factors*, vol. 44, no. 2, pp. 314–334, Summer 2002.

[15] G. Abe and J. Richardson, "The influence of alarm timing on braking response and driver trust in low speed driving," *Safety Sci.*, vol. 43, no. 9, pp. 639–654, Nov. 2005.

[16] E. E. Wiese and J. D. Lee, "Auditory alerts for in-vehicle information systems: The effects of temporal conflict and sound parameters on driver attitudes and performance," *Ergonomics*, vol. 47, no. 9, pp. 965–986, Jul. 2004.

[17] J. Edworthy, E. Hellier, K. Walters, W. Clift-Mathews, and M. Crowther, "Acoustic, semantic and phonetic influences in spoken warning signal words," *Appl. Cogn. Psychol.*, vol. 17, no. 8, pp. 915–933, Dec. 2003.

[18] M. J. Traxler, *Introduction to Psycholinguistics: Understanding Language Science*. Chichester, U.K.: Wiley, 2011, pp. 105–116.

[19] J. L. McClelland and J. L. Elman, "Interactive processes in speech perception: The TRACE model," in *Parallel Distributed Processing*, vol. 2. Cambridge, MA, USA: MIT Press, 1986, pp. 58–121.

[20] C. Wu and Y. Liu, "Queuing network modeling of driver workload and performance," *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 3, pp. 528–537, Sep. 2007.

[21] C. Wu, M. Berman, and Y. Liu, "Optimization in brain?–Modeling human behavior and brain activation patterns with queuing network and reinforcement learning algorithms," in *Computational Neuroscience*. New York, NY, USA: Springer, 2010, pp. 157–179.

[22] C. Wu and Y. Liu, "Queuing network modeling of transcription typing," *ACM Trans. Comput.-Hum. Interact. (TOCHI)*, vol. 15, no. 1, p. 6, May 2008.

[23] G. Zhao, C. Wu, and C. Qiao, "A mathematical model for the prediction of speeding with its validation," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 2, pp. 828–836, Jun. 2013.

[24] C. Wu and Y. Liu, "Queuing network modeling of the psychological refractory period (PRP)," *Psychol. Rev.*, vol. 115, no. 4, pp. 913–954, Oct. 2008.

[25] C.-J. Lin and C. Wu, "Mathematically modelling the effects of pacing, finger strategies and urgency on numerical typing performance with queuing network model human processor," *Ergonomics*, vol. 55, no. 10, pp. 1180–1204, Oct. 2012.

[26] G. Zhao and C. Wu, "Mathematical modeling of driver speed control with individual differences," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 43, no. 5, pp. 1091–1104, Sep. 2013.

[27] C. Wu, Y. Liu, and C. M. Quinn-Walsh, "Queuing network modeling of a real-time psychophysiological index of mental workload—P300 in event-related potential (ERP)," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 38, no. 5, pp. 1068–1084, Sep. 2008.

[28] Q. Ma, J. Jin, and L. Wang, "The neural process of hazard perception and evaluation for warning signal words: Evidence from event-related potentials," *Neurosci. Lett.*, vol. 483, no. 3, pp. 206–210, Oct. 2010.

[29] J. Valls-Solé *et al.*, "Reaction time and acoustic startle in normal human subjects," *Neurosci. Lett.*, vol. 195, no. 2, pp. 97–100, Aug. 1995.

[30] J. Qin and S. Han, "Neurocognitive mechanisms underlying identification of environmental risks," *Neuropsychologia*, vol. 47, no. 2, pp. 397–405, Jan. 2009.

[31] F. Pulvermüller, M. Härle, and F. Hummel, "Walking or talking?: Behavioral and neurophysiological correlates of action verb processing," *Brain Lang.*, vol. 78, no. 2, pp. 143–168, Aug. 2001.

[32] J. N. M. Fuster, "Network memory," *Trends Neurosci.*, vol. 20, no. 10, pp. 451–459, Oct. 1997.

[33] T. D. Blumenthal, "The startle response to acoustic stimuli near startle threshold: Effects of stimulus rise and fall time, duration, and intensity," *Psychophysiology*, vol. 25, no. 5, pp. 607–611, Sep. 1988.

[34] T. D. Blumenthal *et al.*, "Committee report: Guidelines for human startle eyeblink electromyographic studies," *Psychophysiology*, vol. 42, no. 1, pp. 1–15, Jan. 2005.

[35] G. Martino and L. E. Marks, "Perceptual and linguistic interactions in speeded classification: Tests of the semantic coding hypothesis," *Perception*, vol. 28, no. 7, pp. 903–924, 1999.

[36] C. L. Baldwin, "Verbal collision avoidance messages during simulated driving: Perceived urgency, alerting effectiveness and annoyance," *Ergonomics*, vol. 54, no. 4, pp. 328–337, Apr. 2011.

[37] J. A. Foss, J. R. Ison, J. P. Torre, Jr., and S. Wansack, "The acoustic startle response and disruption of aiming: I. Effect of stimulus repetition, intensity, and intensity changes," *Hum. Factors*, vol. 31, no. 3, pp. 307–318, Jun. 1989.

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

[39] P.-S. Jang, "Designing acoustic and non-acoustic parameters of synthesized speech warnings to control perceived urgency," *Int. J. Ind. Ergonom.*, vol. 37, no. 3, pp. 213–223, Mar. 2007.

[40] I. S. Sigalovsky and J. R. Melcher, "Effects of sound level on fMRI activation in human brainstem, thalamic and cortical centers," *Hearing Res.*, vol. 215, no. 1, pp. 67–76, May 2006.

[41] S. Uppenkamp *et al.*, "Locating the initial stages of speech—Sound processing in human temporal cortex," *NeuroImage*, vol. 31, no. 3, pp. 1284–1296, Jul. 2006.

[42] K. B. Laughery, "Computer simulation of short-term memory: A component-decay model," in *The Psychology of Learning and Motivation*, vol. 3, J. T. Spense and G. H. Bower, Eds. San Diego, CA, USA: Academic, 1970, pp. 135–200.

[43] V. Vorhold *et al.*, "The neural basis of risk ratings: Evidence from a functional magnetic resonance imaging (fMRI) study," *Neuropsychologia*, vol. 45, no. 14, pp. 3242–3250, Nov. 2007.

[44] C. Gonzalez, B. A. Lewis, D. M. Roberts, S. M. Pratt, and C. L. Baldwin, "Perceived urgency and annoyance of auditory alerts in a driving context," *Proc. Hum. Factors Ergonom. Soc. Annu. Meet.*, vol. 56, no. 1, pp. 1684–1687, Sep. 2012.

[45] J. Parmanen, "A-weighted sound pressure level as a loudness/annoyance indicator for environmental sounds—Could it be improved?" *Appl. Acoust.*, vol. 68, no. 1, pp. 58–70, Jan. 2007.

[46] C. D. Wickens, *Engineering Psychology and Human Performance*. New York, NY, USA: Harper Collins, 1992.

[47] C. Mulert, J. Gallinat, H. Dorn, W. M. Herrmann, and G. Winterer "The relationship between reaction time, error rate and anterior cingulate cortex activity," *Int. J. Psychophysiol.*, vol. 47, no. 2, pp. 175–183, Feb. 2003.

[48] P. Keuss and M. Van der Molen, "Positive and negative effects of stimulus intensity in auditory reaction tasks: Further studies on immediate arousal," *Acta Psychol.*, vol. 52, no. 1/2, pp. 61–72, Nov. 1982.

[49] C. L. Baldwin, "Cognitive implications of facilitating echoic persistence," *Memory Cognition*, vol. 35, no. 4, pp. 774–780, Jun. 2007.

[50] A. W. Melton, "Implications of short-term memory for a general theory of memory," *J. Verbal Learn. Verbal Behav.*, vol. 2, no. 1, pp. 1–21, 1963.

[51] J. Freedman and W. A. Rumbaugh, "Accuracy and speed of response to different voice types in a cockpit voice warning system," Air Force Inst. Technol., Wright-Patterson AFB, OH, USA, 1983. [Online]. Available: http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA135595.

[52] H. Frenz and M. Lappe, "Absolute travel distance from optic flow," *Vis. Res.*, vol. 45, no. 13, pp. 1679–1692, Jun. 2005.

[53] D. R. Montello, "The perception and cognition of environmental distance: Direct sources of information," in *Spatial Information Theory a Theoretical Basis for GIS*. Berlin, Germany: Springer-Verlag, 1997, pp. 297–311.

[54] F. P. Redlick, M. Jenkin, and L. R. Harris, "Humans can use optic flow to estimate distance of travel," *Vis. Res.*, vol. 41, no. 2, pp. 213–219, Jan. 2001.

[55] B. G. Witmer and P. B. Kline, "Judging perceived and traversed distance in virtual environments," *Presence*, vol. 7, no. 2, pp. 144–167, Apr. 1998.

[56] B. R. Fajen, "Perceptual learning and the visual control of braking," *Perception Psychophys.*, vol. 70, no. 6, pp. 1117–1129, Aug. 2008.

[57] D. Kadihasanoglu, R. D. Beer, and G. P. Bingham, "The dependence of braking strategies on optical variables in an evolved model of visually-guided braking," in *From Animals to Animats 11*. Berlin, Germany: Springer-Verlag, 2010, pp. 555–564.

[58] C. L. Baldwin and J. F. May, "Loudness interacts with semantics in auditory warnings to impact rear-end collisions," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 14, no. 1, pp. 36–42, Jan. 2011.

[59] J. Hawkins, Jr. and S. Stevens, "The masking of pure tones and of speech by white noise," *J. Acoust. Soc. Amer.*, vol. 22, no. 1, pp. 6–13, Jan. 1950.

[60] A. Weber and O. Scharenborg, "Models of spoken word recognition," *Wiley Interdisc. Rev. Cogn. Sci.*, vol. 3, no. 3, pp. 387–401, May 2012.

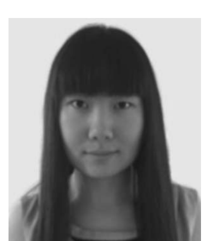
[61] M. G. Gaskell and W. D. Marslen-Wilson, "Integrating form and meaning: A distributed model of speech perception," *Lang. Cogn. Process.*, vol. 12, no. 5–6, pp. 613–656, Oct. 1997.

[62] D. Norris, "Shortlist: A connectionist model of continuous speech recognition," *Cognition*, vol. 52, no. 3, pp. 189–234, Sep. 1994.

[63] O. Scharenborg, "Modeling the use of durational information in human spoken-word recognition," *J. Acoust. Soc. Amer.*, vol. 127, no. 6, pp. 3758–3770, Jun. 2010.

[64] P. A. Luce, "A computational analysis of uniqueness points in auditory word recognition," *Perception Psychophys.*, vol. 39, no. 3, pp. 155–158, Mar. 1986.

[65] L. Ten Bosch, L. Boves, and M. Ernestus, "Towards an end-to-end computational model of speech comprehension: Simulating a lexical decision task," in *Proc. 14th Annu. Conf. Int. Speech Commun. Assoc.*, 2013, pp. 2822–2826. [Online]. Available: http://pubman.mpdl.mpg.de/pubman/faces/viewItemOverviewPage.jsp?itemId=escidoc:1835410.



Yiqi Zhang received the B.S. degree in psychology from Zhejiang University, Hangzhou, China, in 2011. She is currently working toward the Ph.D. degree in industrial and system engineering with The State University of New York at Buffalo, Buffalo, NY, USA. Her current research interest includes studying driver behavior and performance in connected vehicle systems, the development of mathematical models of human response in the interaction with engineering systems, and addressing human cognition with their application in the design of intelligent transportation systems.

Changxu Wu (S'04–M'07) received the M.S. and Ph.D. degrees in industrial and operational engineering from University of Michigan, Ann Arbor, MI, USA, in 2004 and 2007, respectively. Since August 2007, he has been an Associate Professor with the Department of Industrial and Systems Engineering, The State University of New York at Buffalo, Buffalo, NY, USA, where he is also the Director of the Cognitive System Laboratory. He is interested in integrating cognitive science and engineering system design, particularly modeling human cognition system with its applications in system design, improving transportation safety, promoting human performance in human–computer interaction, and inventing innovative sustainable and smart energy systems with human in the loop. He has published more than 36 journal papers in these fields. He is an Associate Editor of IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATIONS SYSTEMS and *Behaviour & Information Technology*.

Jingyan Wan received the B.S. degree in automation and the M.S. degree in control science and engineering from Harbin Institute of Technology, Harbin, China, in 2007 and 2009, respectively. She is currently working toward the Ph.D. degree in industrial engineering with The State University of New York at Buffalo, Buffalo, NY, USA. Her research interests are driver behavior and performance in automated vehicles and ecodriving systems with human in the loop.

AUTHOR QUERY

NO QUERY.

Mathematical Modeling of the Effects of Speech Warning Characteristics on Human Performance and Its Application in Transportation Cyberphysical Systems

Yiqi Zhang, Changxu Wu, *Member, IEEE*, and Jingyan Wan

***Abstract*—Transportation cyberphysical systems (CPSs) aim to improve driving safety by informing drivers of hazards with warnings in advance. The understanding of human responses to speech warnings is essential in the design of transportation CPSs to eliminate hazards and accidents. To date, many works have addressed diverse warning characteristics with experimental approaches. However, the computational model to quantify the effects of warning characteristics on human performance in responses to speech warnings is still missing. Mathematical equations were built to model the effects of lead time, loudness, and signal word choices on human perceptual, cognitive, and motor activities involved in speech warning responses. Different levels of lead time, levels of loudness, and signal word choices served as inputs in the model to predict human error rate and reaction time of speech warning responses. The model was validated with drivers' crash rates and reaction times to speech warnings of upcoming hazards in driving assistant systems in two empirical studies. Results show a good prediction of human performance in responding to speech warnings compared with the empirical data. The application of the model to identify optimal parameter settings in the design of speech warnings in order to achieve greater safety benefits is later discussed.**

***Index Terms*—Human performance modeling, human–computer interaction, intelligent transportation systems.**

I. Introduction

DEATHS and injuries resulting from road traffic accidents has become a major public health problem. According to statistic data published by the National Highway Traffic Safety Administration (NHTSA) in U.S., 5.3 million crashes occurred nationally in 2011 [1]. With regard to improve driving safety, recent advances in Transportation Cyber-Physical Systems (CPS) aim to establish a connected transportation environment by monitoring the status of the physical worlds (e.g., sensors and actuators), connecting it with the cyber worlds (e.g., information, communication, and intelligence), and providing the integrated real-time information among multiple levels, including vehicles to vehicle communication, vehicle to infrastructures communication and in-vehicle information communication [2]. Compared to conventional transportation environment, the connectivity of the transportation CPS allows drivers to learn about the traffic status out of their sight, and provides them with more time to respond to warnings regarding potential hazards [3].

In order to improve the safety of both humans and vehicles, as well as facilitate communication between them, it is important to design warning characteristics based on human performance. While work has been done to increase the communication reliability of connected vehicles, the effectiveness of such systems could not be achieved without drivers making proper and timely responses. Therefore, modeling driver responses to warnings is necessary to achieve effectiveness of warning systems with the human in the loop.

Compared to non-speech auditory warnings, speech warnings are more user-friendly since humans can easily understand and differentiate warnings without specific trainings in memorizing and recognizing warnings [4]. Previous work showed that people working in an operation room had difficulties in recognizing more than half of the non-speech warnings currently in use [5]. Another study indicated that people were unable to distinguish more than six complex warnings [6]. Moreover, previous work found that speech warnings led to a faster reaction time than non-speech warnings regarding spatial information [7]. As a consequence, speech warnings can be widely applied to the Transportation CPS with different warnings in diverse traffic situations.

To date, many empirical studies have examined the influence of warning characteristics on human performance, such as content, perceived hazard, familiarity, signal word, warning sources, and number of items in speech warnings, on human behavior and performance [8]–[11]. Existing empirical has been shown that warning lead time, loudness and signal word choice have significant effects on driver responses to speech warnings. Lead time is defined as the available time for responses from the start of the speech warning until the occurrence of the collision [12]. Studies showed early warnings led to shorter reaction times to collisions than either middle or late warnings [13]–[15]. The warning loudness was found to have a significant effect on urgency expression [16]. In terms of warning

Manuscript received September 14, 2015; revised January 21, 2016; accepted February 27, 2016. This work was supported by the National Science Foundation. The Associate Editor for this paper was Y. Gao.

The authors are with the Department of Industrial and Systems Engineering, The State University of New York at Buffalo, Buffalo, NY 14260 USA (e-mail: yiqizhan@buffalo.edu; seanwu@buffalo.edu; jingyanw@buffalo.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2016.2539975

semantics, the different signal words chosen in speech warnings significantly influence a human's judgment of the urgency level of a situation [17]. However, the behavioral approach used in existing empirical studies to assess the effectiveness of speech warnings can be highly task-dependent, time consuming, and high-cost. The modeling approach we adopted in the current work will provide the predictions of human performance under the different levels of the modeled warning characteristics by running the developed model, and help designers improve their warning designs in Transportation CPS.

To our best knowledge, there are few mathematical models that predict human responses to speech warnings. Two major psycholinguistic models, the COHORT model and TRACE model, have described the mechanism of how human recognize and process spoken words in general. The COHORT model is a bottom-up verbal model that explains the lexical access for spoken word perception [18]. In the stage of activation, perception is influenced by auditory stimulation such that all words matching the perceived acoustic profile are activated, serving as a *cohort*. The selection stage refers to the process of selecting consistent input and eliminating candidate words that no longer match the input. Once the single candidate is isolated from the cohort, word recognition is accomplished. Unlike the COHORT model, the TRACE model is an interactive activated simulation model. The main feature of the model is the ability to describe the interaction of units including within-level inhibition and between-level facilitation [19]. The cascaded activation mode in the TRACE model enables the activation word-level processing units sooner after the activation of the feature-level processing units. The word with the most support from the bottom layers will increase its activation until only one candidate is left standing. These two cognitive models laid the significant foundation on understanding the mechanism of speech perception and processing.

However, the COHORT model and the TRACE model focus on the speech perception and recognition instead of human responses to speech. Therefore, they cannot be used to predict human performance in their responses to speech warnings. Meanwhile, both psycholinguistic models focus on general mechanism of speech processing rather than different characteristics of speech warnings so that they are not able to predict the effects of different characteristics of speech warnings on human responses. Moreover, neither COHORT nor TRACE model is a mathematical model. Mathematical models are indispensable to predict how human respond to speech warnings under the influence of warning characteristics in order to be applied in the design of transportation CPS. Therefore, new models are still needed to model how different characteristics of the speech warnings affect human responses with the mapping between the meaning of speech warnings and the target response actions.

The present work addresses this problem by developing a mathematical model to predict human responses to speech warnings in human–machine systems. This paper extended the model presented in [72] by integrating the algorithm of reinforcement learning in modeling the route choice in the processing of speech warnings and quantifying human reaction error rate and reaction time in speech warning responses. Three main speech-warning parameters are discussed: lead time, loudness, and signal word choice. As the causes of accident in reality can be very complex, the errors in initial responses and the slowed responses to warnings are two of major causes that leaded to traffic accidents. Therefore, accident rate is modeled as the outputs of the model with this two causes being considered and is tested with two empirical studies. In addition, the applications of the model were discussed in setting up the warning parameters to optimize the design of transportation cyber-physical system in terms of human performance. The interface of web-based software was proposed for designers as an easy-to-use technology to design different speech warning parameters associated with human performance.

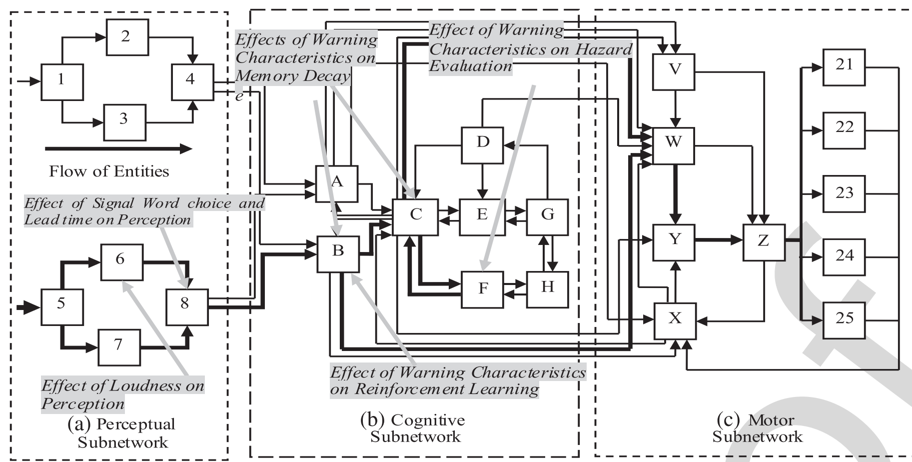
II. Modeling Mechanism and Model Enhancement

A. Overview of Queuing Network-Model Human Processor (QN-MHP)

Queuing Network-Model Human Processor (QN-MHP) is a computational architecture that integrates three discrete serial stages of human information processing (i.e., perceptual, cognitive, and motor processing) into three continuous subnetworks (see in Fig. 1). Each subnetwork is constructed of multiple servers and links among these servers. Each individual server is an abstraction of a brain area with specific functions, and links among servers represent neural pathways among functional brain areas. The neurological processing of stimuli is illustrated in the transformation of entities passing through routes in QN-MHP. Since this architecture was established, QN-MHP has been applied to quantify various aspects of human cognition and performance, such as human mental workload [20], and the reinforcement learning process [21]. In terms of the perceptual subnetwork, new equations have been integrated to model eye movements, and speed perception [22], [23]. The cognitive subnetwork has been improved to model textual information chunking [26], inhibition incompatible responses and choice reactions [24], dual task interference [25], and complex decision making [26]. Moreover, applications of QN-MHP indicate its success in modeling motor program retrieval [26], error corrections [25], bimanual coordination in typing tasks, and driver speed control [23], [26].

B. Enhancements of Queuing Network-Model Human Processor (QN-MHP)

In the present work, the mathematical model was proposed based on architecture of QN-MHP to predict human performance in speech warning responses with system operation tasks (e.g., driving a vehicle) based on neurological findings [34]–[38], [42]–[49]. Although several mathematical models based on the QN-MHP have been successfully built to predict driver behaviors such as speed and lateral control, the model to predict human responses to speech warning is still missing. The highlighted servers with labels in Fig. 1 illustrated the servers to be enhanced with the equations developed in the current work and the processing of speech warnings with the "Flow of Entities."



Perceptual Subnetwork	Cognitive Subnetwork	Motor Subnetwork
1. Common visual processing	A. Visuospatial sketchpad	V. Sensorimotor integration
2. Visual recognition	B. Phonological loop	W. Motor program retrieval
3. Visual location	C. Central executive	X. Feedback information collection
4. Visual recognition and location integration	D. Long-term procedural memory	Y. Motor program assembling and error detecting
5. Common auditory processing	E. Performance monitor	Z. Sending information to body parts
6. Auditory recognition	F. Complex cognitive function	21-25: Body parts: eye, mouth, left hand, right hand, foot
7. Auditory location	G. Goal initiation	
8.Auditory recognition and location integration	H. Long-term declarative & spatial memory	

Fig. 1. The enhanced servers of the QN-MHP architecture with the equations to be developed in the current work, and the general structure of QN-MHP (developed in [20], [21], [24], and [27]; and all of the published mathematical equations in QN-MHP can be found at: http://www.acsu.buffalo.edu/~seanwu/QNMHPMath/MathModelQNMHP_Online.htm).

In the speech warnings response task, the stimuli of speech warnings entered into the auditory perceptual subnetwork. The stimuli firstly arrived at Server 5, representing the middle and inner ear (common auditory processing). The parallel auditory pathways transmitted the auditory information through the neuron pathways from the dorsal/ventral cochlear nuclei to the inferior colliculus presented by Server 6 (auditory recognition) and from the ventral cochlear nucleus to the superior olivary complex represented by Server 7 (auditory location).

Then the auditory information was integrated at Server 8, representing the primary auditory cortex and the planum temporale (auditory recognition and location integration. The speech warnings with specific loudness and semantic features were then transmitted to the left-hemisphere posterior parietal cortex presented as Server B (phonological loop).

A route choice located at Server B with a shorter route directly connecting to Server W (motor programs retrieval) representing basal ganglia, and a longer route connecting to Server C (central executive) and Server F (complex cognitive function), and eventually leading to Server W. The shorter route represented a processing in emergent situations and the longer route involved the stage of hazard evaluation in less emergent situations. Those motor programs at Server W were then assembled at Server Y (motor program assembling and error detecting) and initialized at Server Z representing primary motor cortex, sending out the neural signals to body parts (Servers 21–25).

1) Modeling the Effect of Speech Warning Parameters on the Probability of Route Choice in Reinforcement Learning: The modeled routes in QN-MHP were presented in Fig. 1. As it showed at Server B, entities could choose one of the two routes to move to either Server C (long route) or Server W (short route). The division of the two routes was modeled with the route choice at Server B. Previous fMRI studies indicate two stages involved in processing warning signal words associated with hazards [28]. One stage is a rapid automatic activity and the other stage involves the activation of the hazard evaluation. The rapid automatic activity with a shorter response time to warnings could be represented by the shorter route (Route I) of warning responses learned through experiences in urgent situations [29], [30]. The other activity involving a hazard evaluation process could be represented by the longer route (Route II) of warning responses learned through experiences in non-urgent situations [35]. To process information with Route II, the human would take a longer time to respond as more servers were involved in this route. In the meantime, the human would have a lower error rate of responses since entities were processed through critical servers (Servers C and F) could correct errors to a certain degree.

The probability of choosing a route could be the result of learning from the connections of warning characteristics and associated hazards in daily life. Previous fMRI studies showed that people learned responses to auditory stimuli with a co-activation of the motor/premotor cortex and the primary auditory cortex [31]. As the neuron in motor and premotor cortex (Server W) fired repeatedly when the human processed associated warnings, the correlation of neuronal firing of connected cortical cells was translated into their connection strength [32]. At the beginning of the learning, entities of speech warnings with different loudness levels or signal words might have equal chances to enter either route. Then the probability of route choosing would be updated as humans learned from association between specific loudness levels/signal words and urgency of hazards.

Whether a situation was considered to be an emergency was determined by certain criteria of loudness levels and signal words. In terms of warning loudness, Blumenthal [33] reported that a 50% probability threshold of a startle response was 85 dB. Studies have shown the increasing of the acoustic stimuli intensity leads to an increase in response magnitude and amplitude, and a decrease in response onset latency [34]. For signal word choices, different signal words expressed different perceived urgency levels (Hollander & Wogalter, 2000). Therefore, speech warnings with its loudness higher than 85 dB or a particular signal word (e.g. “Danger”) would represent an emergency situation.

Moreover, the incompatibility of warning loudness and word semantics indicating different hazard urgency levels took longer time for human to respond [28]. This incompatibility might result in entities traveling through a longer route (Route II) with higher chance in order to solve the incompatibility problem [35]. The probabilities of choosing route I (p_I) and route II (p_{II}) for speech warnings with certain loudness levels and signal words were obtained from the simulation results (see Q online learning algorithms in the Appendix).

2) Modeling the Effect of Speech Warning Characteristics on the Warning Perception, Memory Decay and Hazard Evaluation: The choices of servers and where to integrate equations were determined by the brain area are influenced by warning characteristics. Studies suggested loudness and signal word choice have significant effects on human behaviors [36]–[39]. It has been shown that the activation of lower auditory processing level increased with the sound level increased [40]. Therefore, the effect of loudness on speech warning perception was modeled at Server 6. The semantic features of signal words are recognized at the superior temporal sulcus, which was modeled at Server 8 [41].

Due to the interference caused by the speech warnings on the on-going tasks, memory decay may occur [42]. The effect of warning lead time on memory decay was modeled in the working memory system regarding auditory processing represented by Servers B and C. Previous fMRI studies indicated that hazard evaluation activated the medial prefrontal cortex, the inferior frontal gyrus, the cerebellum, and the amygdale [43], which were presented by Server F.

III. Mathematical Formulation of Modeling Mechanisms and the Enhancement of the QN-MHP

A. Modeling the Effects of Loudness and Signal Word Choice on Perceived Urgency and Annoyance of Speech Warnings

1) Modeling the Relationship Between Loudness and Perceived Urgency/Annoyance: The relations between changes in loudness and changes in perceived urgency can be quantified by the Stevens Power Law [38]. The loudness was reported having a positive relationship with urgency expression [44]. Therefore, the perceived urgency (U_L) and annoyance (A_L) as a function of warning loudness was modeled by the following equations:

$$\log(U_L) = m_U \log(L) + k_U + \varepsilon_1 \tag{1}$$

$$\log(A_L) = m_A \log(L) + k_A + \varepsilon_2 \tag{2}$$

where L denotes the loudness level and m and k quantify the relationship between perceived value and objective loudness change. The relationship between intensity and perceived urgency/annoyance was quantified [44]. The Stevens' power law states that the loudness (L) is proportional to $I^{0.3}$, where I is the sound intensity [45]. Therefore, the parameters are quantified as: $m_U = 1.33$, $m_A = 1.45$, $k_U = -0.64$, $k_A = -0.91$. ε_1 and ε_2 are normally distributed random factors following distribution [0, 0.7] and [0, 0.86], respectively [36].

2) The Relationship Between Signal Word Choice and Perceived Urgency: Considerable research efforts have been indicating a stable relationship between signal word choice and perceived urgency. Hollander & Wogalter (2000) reported ratings in carefulness expressed in a descending order by the following five signal words: deadly, danger, warning, caution and notice. Other studies have found similar results. These words covered a wide range of urgency ratings and have been studied before in detail (Barzegar & Wogalter, 1998; Hollander & Wogalter, 2000) using the word “notice” rather than “note.” The perceived urgency of “danger,” “caution,” and “notice” spoken by a female voice on a 100 points scale are quantified as 90.53,72.40, and 46.81 [44].

B. Modeling the Error Rate in Speech Warning Responses

Speech warning parameters have different influence on speech warning response error rate in different stages of speech warning responses. When humans processed speech warnings through route I, the error rate was mainly influenced by the effects of loudness and signal words on speech warning perception. When speech warnings were processed through route II, the error rate in the speech warning responses was also influenced by the effects of lead time on potential memory decay of the speech warnings and hazard evaluation.

1) Modeling the Effect of Loudness and Signal Word Choice on Error Rate: Errors in speech warning responses could result from the shortcoming of perception, memory, cognition and the failure in motor execution [46]. Errors in speech warning responses include no responses to correct warnings (e.g., failures in recognizing speech warnings and misjudging hazards associated with warnings) or incorrect responses to warnings (e.g., accelerating instead of braking towards a forward collision). The error rate (I_E) is modeled as a function of the speech warning loudness and signal word choice and the corresponding probability of route choices. A warning with higher urgency is correlated with higher arousal strength, which may result in a startle reflex and lead to a higher chance of poorly processing the warning signal words [28]. This autonomic activity can be represented as entities traveling through route I with a higher chance of making errors such that entities are not processed in critical Server C and Server F. Both loudness and semantic features relevant to the expressed urgency of the speech warnings have influence on error rate in the perception of speech warnings [47]. Also, a positive correlation between loudness and error rate was found in an empirical study [48]. The error rate in route I is then modeled with a positive correlation with perceived urgency expressed by word loudness and word semantics.

The entity processed through route II involves the central executive and hazard evaluations at Servers C and F. The effect of loudness on error in response would decrease after the entity passed the phonological loop due to the decay of the echoic memory [52]. Further processing of the entity led to pattern recognition or semantic analysis of the speech warnings (at Server C) and the corresponding hazard was evaluated in the decision making stage (at Server F) [28], [49]. Therefore, the error rate in route II was modeled with a correction of errors brought in by the loudness and semantic properties of the speech warnings in the perceptive stage of speech warnings.

In summary, the error rate $(I_{E,i})$ of route i $(i = I$ or $II)$ is modeled with the following equation (3) with the perceived urgency (U_L) and annoyance (A_L) of speech warnings due to different loudness levels, and the perceived urgency of speech warnings due to different signal words (U_S). Since there is no difference of perceived annoyance due to different signal words (A_S), it is not inputted in modeling the error rate

$$I_{E,i} = \begin{cases} (U_L + U_S) \times 0.5, & i = I \\ (U_L - A_L) \times 0.5, & i = II \end{cases} \tag{3}$$

where L is the speech warnings loudness and S is the signal words. U_L and A_L are the perceived urgency and annoyance of warning loudness obtained from (1) and (2); U_S is the urgency of signal word choice. According to the perceived urgency for signal word scales, the perceived urgency for word semantics (U_S) is 0.90, 0.72 or 0.47 for signal words “Danger,” “Caution,” “Notice,” respectively [44].

The overall error rate in the responses to speech warnings is then modeled by adding up the error rate with the probability in each route. The effect of speech warning parameters on route choice error rate (I_E) can be modeled as the combined effect of the speech warning loudness and signal word choice:

$$I_E = \sum_{i=1}^{2} I_{E,i} \times p_i \tag{4}$$

where $I_{E,i}$ denotes the error rate when a speech warning travels through route i. p_i denotes the probability of information processing through route i.

Then the equation (4) for the effect of speech warning loudness and signal word choice on error rate (I_E) is updated by the following general equation:

$$I_E = (L^{m_U} \times 10^{k_U - 2} + U_S) \times 0.5 \times p_I + \left((L^{m_U} \times 10^{k_U - 2} - L^{m_A} \times 10^{k_A - 2}\right) \times 0.5 \times p_{II} \tag{5}$$

where L denotes the loudness level in dB. U_S is the perceived urgency level with different signal word choice. p_I and p_{II} are probabilities of choosing route I (the shorter route) and route II (the longer route) respectively obtained from the simulation results of the reinforcement learning in Appendix. m_U and k_U are parameters to quantify the power law of perceived urgency and loudness. m_A and k_A are parameters to quantify the power law of perceived annoyance and loudness.

2) Modeling the Impact of Lead Time on Error Rate: Drivers tend to respond to the speech warning when the corresponding hazard is within sight [13]. When there is a relatively long lead time before the actual hazard occurrence, the human may perform normal operations and monitor the situation. Therefore, the memory of the speech warnings may decay and the corresponding accuracy rate of upcoming hazard estimation may increase the error rate in responses to speech warnings.

The probability of information retrieving (p) is modeled as a function of time (t) starting from the information presented to humans in [42] as follows:

$$p = e^{\text{at}}, [42] \tag{6}$$

where $a = -0.02$ based on parameter settings of MHP [50].

In the proposed speech warning responses model, the effect of lead time on memory decay (I_{MD}) is computed at Servers B and C in QN-MHP, representing the working memory system regarding auditory information processing

$$I_{MD} = \frac{1}{e^{-0.02t}\text{lead}}. \tag{7}$$

In the above equation, t_{lead} denotes the lead time for speech warning responses.

In terms of hazard estimation, a human will react to speech warnings when a perceived hazard reaches a certain threshold. The effect of hazard evaluation accuracy on error rate (I_H) can be modeled by the difference between the perceived value and the actual value of the hazard in the following equation:

$$I_H = \frac{H_p}{H_0} \tag{8}$$

where H_p denotes the perceived value of hazard and H_0 denotes the actual value of hazard.

In summary, the error rate (r) in speech warning responses is extended by adding the effects of loudness and signal word choice modeled in (5), and the effect of lead time modeled in (7) and (8) as follows:

$$r = I_E + I_{MD} \times I_H + \varepsilon_3 \tag{9}$$

where I_E denotes the error from signal word perception and recognition under the effect of speech warning loudness and signal word choice, I_{MD} denotes the error from memory decay, I_H denotes the error from hazard location estimation. ε_3 is a random factor following normal distribution [0, 0.1] [51].

C. Modeling the Reaction Time in Speech Warning Responses

The reaction time was defined as the time duration from the time the speech warning occurs to the time the human starts to react. As assumed in QN-MHP, entity processing time at an individual server is independent of arrivals of entities, and routing is independent of the state of the system. Therefore, the reaction time of a speech auditory stimulus can be modeled by summarizing the processing time of all the servers on the route.

Consequently, the reaction time (RT_i) to speech warnings through route i is modeled as:

$$RT_i = \begin{cases} T_5 + T_6 + T_8 + T_B + T_W + T_Y + T_Z, & i = I \\ T_5 + T_6 + T_8 + T_B + T_C + T_F + T_C & \\ \quad + T_W + T_Y + T_Z & i = II \end{cases} \tag{10}$$

where T_k is the processing time of auditory stimulus at Server k. The processing time of servers in perceptual, cognitive, and motor subnetwork are 42 ms, 24 ms, and 18 ms [24].

The effect of loudness on reaction time is modeled in the initial processing of auditory stimuli in Server 6

$$T_6 = \frac{T_{6(0)}}{U_L} \tag{11}$$

where $T_{6(0)}$ is the initial entity processing time in Server 6 and U_L denotes the effect of loudness on perceived urgency.

The effect of signal word choice on reaction time can be modeled by the following equation:

$$T_8 = \frac{T_{8(0)} \times n_i}{U_s} \tag{12}$$

where $T_{8(0)}$ is the entity processing time in Server 8 and n_i is the number of words in the ith speech warning. U_s denotes the urgency level expressed by the initial words (e.g., signal words) in the speech warnings.

All in all, the equation (10) for modeling reaction time of speech warnings through route i is updated as:

$$\begin{aligned} RT = &\left(T_5 + \frac{T_{6(0)}}{U_L} + \frac{T_{8(0)}}{U_s} + T_B + T_W + T_Y + T_Z\right) \\ &\times p_I + \left(T_5 + \frac{T_{6(0)}}{U_L} + \frac{T_{8(0)}}{U_s} + T_B + T_C + T_F \right. \\ &\left. + T_C + T_W + T_Y + T_Z\right) \times p_{II} + \varepsilon_4. \end{aligned} \tag{13}$$

In the above, T_k denotes the processing time of the auditory stimulus at Server k $(k = 5-8, B, C, F, W-Z)$. U_L is the perceived urgency level with different levels of loudness. p_I and p_{II} are probabilities of choosing route I (the shorter route) and route II (the longer route), respectively. ε_4 is a normally distributed random factor following distribution [0, 0.3] [13].

D. The Application of Speech Warning Response Model in Driving and Warning Responses

The following section presents the application of the speech warning responses model in modeling human responses to speech warnings in Transportation CPS systems (e.g., in-vehicle information systems and connected vehicle communication systems). Warning responses in a driving task include the releasing of the accelerator pedal when drivers are accelerating and the change in braking pedal when drivers are already braking (i.e., foot on brake pedal) or on their way to brake (i.e., releasing the accelerator). The parameters of speech warnings are loudness and signal word choice, as well as lead time. The drivers tend to respond once the speech warning begins when they hear the signal words (e.g. "Notice," "Caution," "Warning," and "Danger"). QN-MHP was used to estimate the reaction to the speech warnings starting from perceiving the information from speech warnings to transmit neural signals to the foot server (Server 25).

1) The Hazard Evaluation in the Driving and Speech Warning Responses Tasks: When the speech warnings are presented to a driver, he/she will continuously evaluate the potential hazard based on the information obtained from visual perception and from speech warnings (e.g., estimated distance). Previous work studied the effects of motion factors (e.g., optical flow rate, optical density of texture and edge rage) and cognitive factors (e.g., perceived time, actual speed) on the traversed distance estimation [52]–[54]. Traveling speed had a significant effect on distance estimation, with slower speed resulting in more accurate distance estimation. The relationship between actual distance and estimated traversed distance (D_P) was quantified with Steven's power law [55]

$$D_P = D_0^{b^v} \tag{14}$$

where D_0 denotes the actual distance between the current position of warning receiving vehicle and the potential hazard location when speech warning is presented, while v denotes the instant speed $(b = 0.955)$ [55]. Based on the definition, the actual distance D_0 is modeled as:

$$D_0 = v_0 t_{\text{lead}} + \frac{1}{2} a_0 t_{\text{lead}}^2. \tag{15}$$

When the perceived distance is shorter than the minimum safety headway, drivers may react to the speech warnings directly. Otherwise, drivers continue to drive and react to speech warnings until perceived distance (D_p) reaches the threshold $(D_p = D_h)$. The hazard evaluation effect on crash rate is modeled as

$$I_H = \frac{D_h}{D_0} = D_0^{b^{v(t)}-1}. \tag{16}$$

The instant speed (v) and acceleration (a_t) at time t is modeled in [23] as follows:

$$v(t) = v_0 + a_t(\Delta t), [23] \tag{17}$$

where v_0 denotes the initial speed and a_t denotes the acceleration at time t.

The constant rate of deceleration $(a_t(\Delta t))$ is modeled in [56] as follows:

$$a_t(\Delta t) = \frac{k}{2} \times \phi \times \frac{\dot{\theta}}{\theta}, [56] \tag{18}$$

where ϕ is the global optic flow rate of the textured ground surface, a proportion of speed as long as eye height is constant. The global optic flow rate is constant in a braking task. The ratio $\dot{\theta}/\theta$, where θ and $\dot{\theta}$ are the optical angle and rate of expansion of approached object, respectively, is approximately equal to v/S. Therefore, the ideal deceleration can be expressed in terms of the optical variable by substituting ϕ for v and $\dot{\theta}/\theta$ for

v/S. Novices tended to initiate emergency braking earlier than necessary when initial speed was slow and to a lesser extent, which brought in a parameter k of driving experiences ($0 < k < 1$). The parameter k is quantified by the annual mileage divided by a maximum value of annual mileage in general.

The ratio of the object's optical angle to rate of expansion of approached object ($\dot{\theta}/\theta$) specifies the time-to-collision (TTC) with the object as long as the current velocity is held constant. The ratio is modeled in [57] as follows:

$$\frac{\dot{\theta}}{\theta} = \mathrm{TTC}, [57]. \tag{19}$$

The perceived time-to-collision (TTC_p) will be affected by the existence of the lead vehicle. TTC is the actual time to collision that the vehicle will be able to avoid a collision without exceeding the assumed maximum deceleration, which is represented as t_{lead} as above

$$\frac{\dot{\theta}}{\theta} = \mathrm{TTC}_p = t_{\mathrm{lead}} \times \exp(LV). \tag{20}$$

In the above, LV is a dichotomous variable of the lead vehicle in order to model the effect of the lead vehicle on TTC_p ($0 =$ without lead vehicle; $1 =$ with lead vehicle).

In summary, the effect of hazard evaluation on crash rate is modeled as:

$$\begin{aligned} I_H &= \frac{D_p}{D_0} = D_0^{b^{v_0 + \frac{k}{2} \times \phi \times \mathrm{TTC}_p} - 1} \\ &= \left(v_0 t_{\mathrm{lead}} + \frac{1}{2} a_0 t_{\mathrm{lead}}^2\right)^{b^{v_0 + \frac{k}{2} \times \phi \times t_{\mathrm{lead}} \times \exp(LV)} - 1}. \end{aligned} \tag{21}$$

2) Modeling the Crash Rate in Speech Warning Responses: The modeling of crash rate has to consider the additional impact of warning lead time. Even if the driver makes correct responses, lack of time to respond will also result in accidents. When the lead time is shorter than the minimum brake-to-maximum response time ($t_{\min}$), the drivers may not avoid the collision even when they correctly respond immediately. Therefore an effect of lead time on crash rate is modeled as:

$$t_{\min} = \frac{v_0}{a_{\mathrm{average}}} + RT = \frac{v_0}{\frac{1}{2}|a_0 + a_{\max}|} + RT \tag{22}$$

$$I_{LT} = \frac{t_{\min}}{t_{\mathrm{lead}}}. \tag{23}$$

The impact of parameters (i.e., loudness and signal word choice) of speech warning on crash rate (R_{crash}) can be modeled by combining Equations (5), (21), (23) as follows:

$$R_{\mathrm{crash}} = I_E + I_{MD} \times I_H \times I_{LT} + \varepsilon_5 \tag{24}$$

where I_E denotes the error from signal word perception and recognition under the effect of speech warning loudness and signal word choice, I_{MD} denotes the error from memory decay, I_H denotes the error from hazard location estimation. I_{LT} denotes the effect of lead time on crash rate. ε_5 is a normally distributed random factor following normal distribution [0, 0.05] [14].

IV. The Validation of the Speech Warning Response Model

In order to validate the speech warning responses model, the following section provides the prediction results of two experimental studies in terms of driver responses to speech warnings. The first study conducted by our research group studied the effect of lead time on driver responses to speech warnings. In order to validate the model, the model predictions for response time and crash rate are shown and compared to experimental data. The second study from a published work examined the effect of loudness and signal word choice of warnings on rear-end collision [58]. Due to a lack of detailed information in the second study, the lead time and hazard evaluation was assumed to have no additional effect on modeling crash rate. The model predictions for crash rate and subjective ratings for perceived urgency and annoyance are shown and compared to experimental data. To validate the speech warning response model, the comparability of model predictions and experimental results were quantified by the Pearson correlation coefficient (R squared) as well as the root mean-squared error (RMSE).

A. Experiment 1

The first experiment involving a driving simulator was conducted to study the impact of lead time on human responses to speech warnings.

1) Participants: Thirty-two participants (24 males, 8 females),draftrules with ages ranging from 18 to 26 years participated in the study. All of them were licensed drivers and had normal or corrected-to-normal vision. None of the drivers had previously participated in any simulator or crash avoidance studies.

2) Apparatus: A STISIM driving simulator (STISIMDRIVE M100K, Systems Technology Inc, Hawthorne, CA) was used in the study. It comprises a Logitech Momo steering wheel with force feedback (Logitech Inc, Fremont, CA), a throttle pedal, and a brake pedal. The STISIM simulator was installed on a Dell Workstation with a 256 MB PCIe $\times$ 16 nVidia graphics card, Sound Blaster X-Fi system, and Dell A225 Stereo System. Driving scenarios were presented on a 27-inch LCD with 1920×1200 pixel resolution. A speaker in front of the participant provided auditory messages in a digitized human female voice with a speech rate of ~150 words/min and loudness level of ~70 dB. Another speaker provided driving sound effects with a loudness level of ~55 dB.

The behavioral measures (time elapsed (s), speed (m/s), acceleration (m/s^2), and distance to the initial location where the scenario starts (m) were automatically collected from the driving simulator and outputted to another identical Dell Workstation. This computer calculates the time to collision (TTC) in real time based on the vehicle's speed and acceleration. When the calculated TTC reached the designed value, the warning would be issued.

3) Scenario Setting: The speech warning would sound before the appearance of the hazard. Each speech warning started with a signal word "Caution" and followed by a description of the collision scenario presented (e.g., A vehicle at your front-left is running red light). The collision scenario description

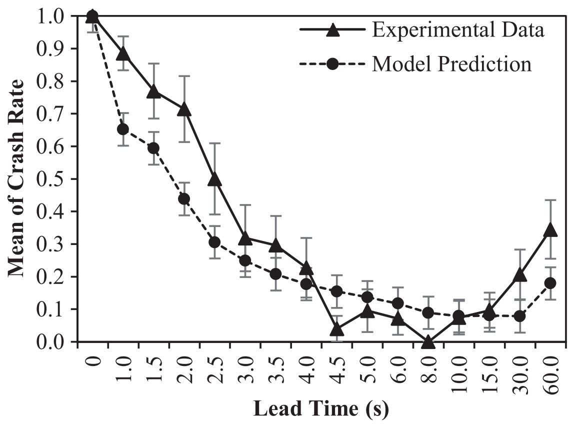


Fig. 2. The comparison of model prediction of crash rate with experiment 1 data (Error bars: +/−1 SE).

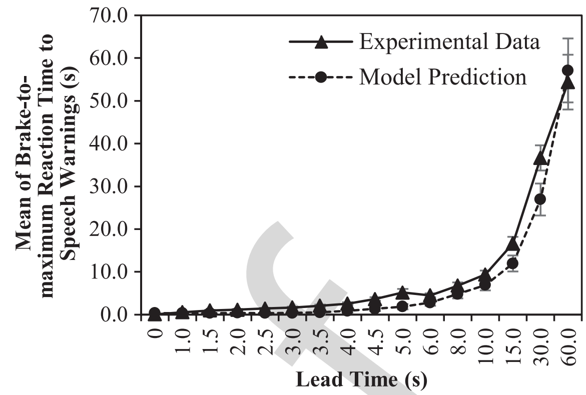


Fig. 3. The prediction of brake-to-maximum response time to the speech warnings (Error bars: +/−1 SE).

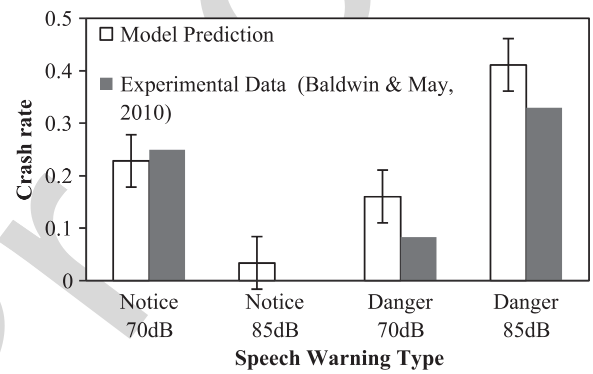


Fig. 4. The comparison of model prediction of crash rate with experiment data (no data of standard error reported in the experiment).

comprised the hazard location and event, which provided the driver with specific information to eliminate any confusion.

The test block used a two-lane (in each direction) urban environment with traffic lights and road signs. There were running vehicles moving in each direction. Speed limit signs with a constant speed limit of 45 mph (20 m/s) were displayed 200 feet (61 m) in front of the driver. Participants were instructed to adjust their speed within the range from 40 mph (18 m/s) to 50 mph (22 m/s) as if they were driving a vehicle in the real world. Sixteen collision scenarios were designed and programmed. A lead vehicle would run at the same speed as the subject vehicle. In order to investigate drivers' responses to speech warnings, their sights of the collision scenario were blocked by other vehicles, and participants could only rely on the warnings to learn about the upcoming collision events.

4) Experiment Design: The current experiment adopted a one-factor experiment design with lead time as an independent variable and collision rate and brake-to-maximum response time as dependent variables. The lead time had 16 levels (0 s, 1 s, 1.5 s, 2 s, 2.5 s, 3 s, 3.5 s, 4 s, 4.5 s, 5 s, 6 s, 8 s, 10 s, 15 s, 30 s, and 60 s). When the lead time was 0, the warning sounded at the same time when the collision event happened. Each subject would go through sixteen collision events with sixteen levels of lead time assigned to each event. The orders of levels of lead time and events were randomized. The normal messages were randomly assigned during the experiment, as long as they did not cause interference with the broadcasting of speech warnings.

The first dependent variable was collision, which specified whether there was collision between a subject's vehicle and a hazard vehicle. The collision rate was then calculated as the percentage of collisions for each level of lead time. Brake-to-maximum response time represented the time period from the present of warnings until drivers reaching the maximum deceleration in the braking responses.

5) Results: The model prediction for crash rate with speech warnings of different lead time levels is shown in Fig. 2. The RMSE was 0.13 with an R square of 0.94. For the brake-to-maximum response time to the speech warnings, Fig. 3 showed the model prediction comparing the experimental results had an R square of 0.97 and RMSE of 3.17.

B. Experiment 2 (Baldwin & May, [64])

The second experimental study examined the effect of loudness and signal word choice of in-vehicle collision warnings on driver responses [64]. Thirty participants were recruited to drive through five different scenarios containing five different hazard events. Speech warnings consisted of the signal word "Notice" or "Danger" presented at either 70 or 85 dBA. The driving sound effects were presented with a loudness level of 55 dB. The crash rate with different warnings and subjective rating of perceived urgency and annoyance were reported.

Due to a lack of detailed information regarding collision event scenario and driver responses, the lead time was set up to be long enough for effective responses in this study since there is no lead time reported ($I_{LT} = 1$). The model prediction for crash rate with different speech warnings is shown in Fig. 4. The RMSE was 0.06 with an R square of 0.90. Fig. 5 shows the model prediction of rating of urgency and annoyance for signal word. The R square of perceived urgency prediction is 1.00 with RMSE of 1.49. The R square of annoyance is not calculated since there is no differences among annoyance ratings of signal words [42].

V. The Application of Prediction of Human Performance in Developing Speech Warnings

Speech warnings in Transportation Cyber-Physical Systems are designed to improve driver safety by providing information about upcoming hazards in an appropriate way so as to

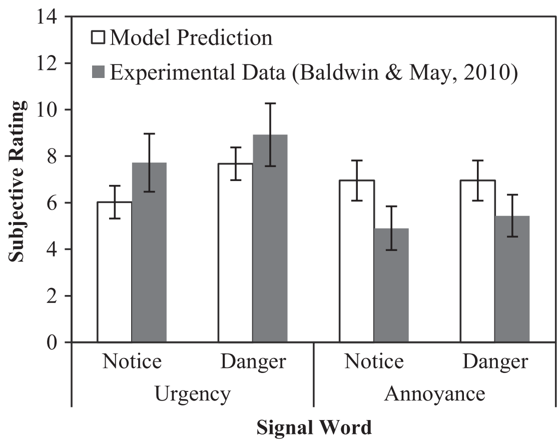


Fig. 5. The prediction of rating of urgency and annoyance for signal words (Error bar: standard deviation).

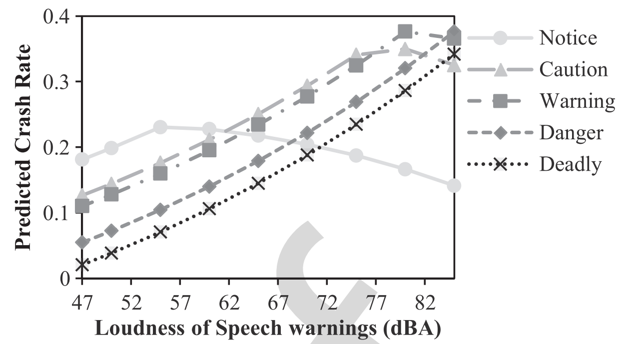


Fig. 6. Predicted crash rate with speech warnings presented at different loudness levels with different signal words.

give drivers enough time to respond. Previous work mainly studied speech warning characteristics through experimental approaches [19], [64]. The developed model in the current work makes it easier for designers to obtain the effects of different speech warning parameters associated with human performance. In particular, the warning lead time, loudness and signal word choice can be optimized by applying the developed model to simulate human performance. Taking the intelligent transportation system as an example, the crash rate will serve as the objective index of potential safety benefit of the speech warnings.

Based on abovementioned modeling results, the model predicted that crash rate would vary with different combinations of lead time, loudness and signal words. Equation (21) is applied to quantify collision rate under different combinations of loudness and signal words with the common noise loudness level of 55 dBA. The threshold of intelligibility (TI) is at the loudness level of 47 dBA, which was defined as the "level at which the listener is just able to obtain without perceptible effort the meaning of almost every sentence and phrase of the connected discourse" [65]. A human, therefore, will not fully recognize and understand warnings with loudness levels below this threshold. The predicted impact of loudness and signal words on crash rate shown in Fig. 6 illustrate the loudness level with range from 47 to 85 dBA with a lead time of zero as an example. The best loudness level to present the signal word "Notice" is 85 dBA, whereas the best loudness level for other signal words is 47 dBA. It is implied that the combination of speech warnings with an intermediate urgency level brought the most safety benefits.

The joint effect of lead time and warning loudness level is shown with the signal word "Caution" (see Fig. 7) as an example. Likewise, the joint effect of lead time and warning signal words is shown with the loudness level of 70 dBA (see Fig. 8). The predicted crash rate has a descending trend as a function of lead time regardless of the impact of loudness level and signal words. Generally speaking, it suggested that early warnings resulted in lower crash rates than did late warnings. As it is shown in Fig. 8, an abrupt decrease of collision rate appeared with longer lead time when the warning was relatively late; the rate of such decrease tended to slow down when the warning was relatively early. The differences in crash rate between different loudness levels and signal words reduced when the lead time was longer. In other words, the impact of loudness and the signal word choice on human responses will decay with the processing of the speech warnings.

Future software can be designed based on the developed models in this work to specify the loudness and signal word choice of speech warnings in the Transportation Cycle-Physical Systems. A sample interface is shown in Fig. 9. With the loudness, signal words and number of words in the speech warnings inputting into the software, the designers of the warning system will be able to obtain the objective parameters regarding human responses, including the predicted crash rate and brake-to-maximum warning response time. Moreover, the subjective rating of the speech warnings could also be obtained by applying this model.

VI. Discussion

In this modeling work, mathematical equations were built within the framework of the Queuing Network Model Human Processor (QN-MHP) to predict human performance in speech warning responses, including human error rate and response time with different warning characteristics. No free parameters were used in the parameter setting. The validation of the model with two laboratory studies indicated its relatively good ability to predict performances in speech warning response with high correlations with behavioral data from two experiments [64].

This work is one of a few mathematical models with analytic solutions in the field of human speech processing. Previous modeling work has explored theories that account for the experimental data of word recognition and speech comprehension [66, 67]. In the review of word recognition models, most modeling work focuses on the mechanism of speech recognition with either verbal models (e.g., COHORT) or simulation models with descriptions of theory implemented in computer programs (e.g., TRACE) [22], [68, 69]. Compared to verbal or simulation models, the conciseness and rigor of mathematical models allows an easier implementation for different systems regardless of the computer language used in the system.

More importantly, few computational models focused on the prediction of human performance in speech warning responses

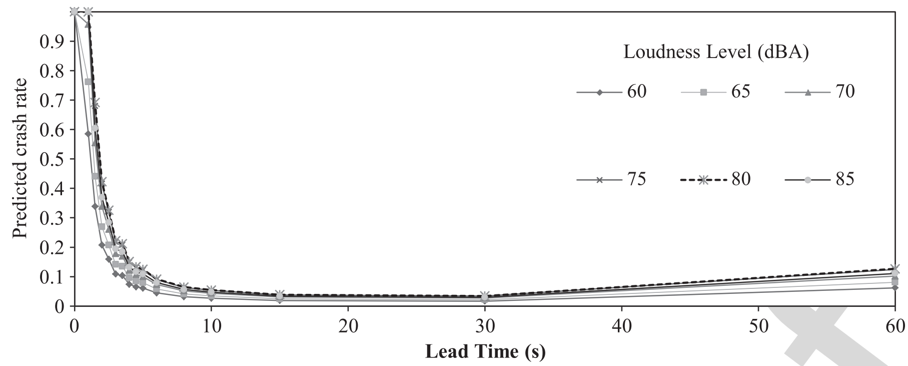


Fig. 7. Predicted crash rate with speech warnings presented at different lead time level and loudness levels (using signal words “Caution”).

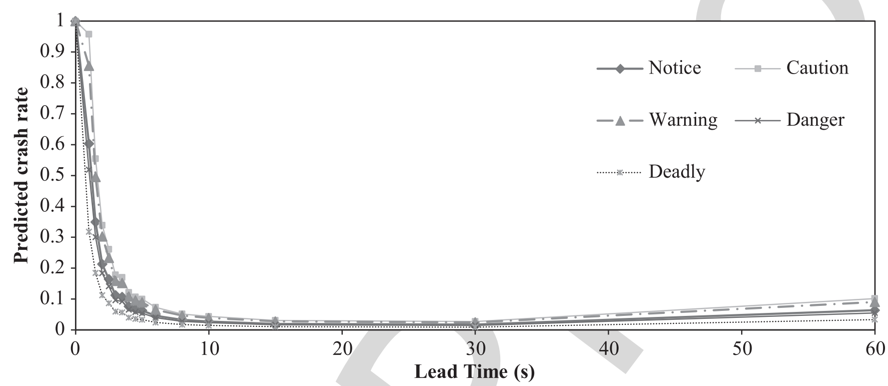


Fig. 8. Predicted crash rate with speech warnings presented at different lead time level and different signal words (at loudness levels = 70 dBA).

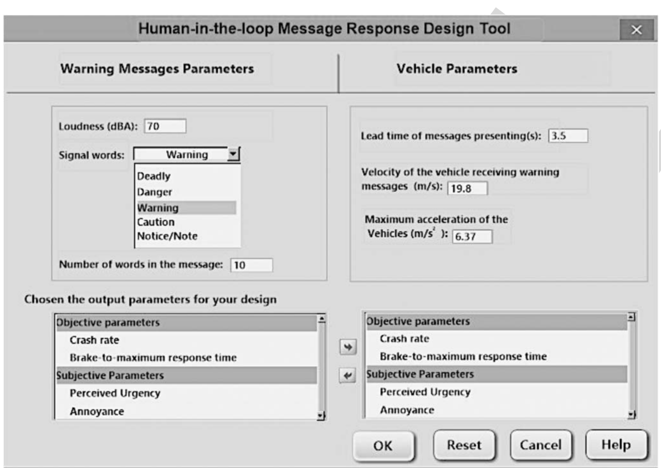


Fig. 9. The sample interface of the software with the application of the model.

and considered the characteristics of speech warnings. The neighborhood activation model focused on the prediction of the unique time point of word recognition [70]. The models that do model human performance (e.g., [71]) haven’t predicted response error rate. In the current work, humans were responding to warnings associated with driving tasks rather than that to isolated words. In this case, the modeled process involves the hazard evaluation associated speech warnings and the selection of proper manual responses with the effects of warning characteristics being modeled. This different emphasis on human response modeling is important in the design of transportation CPS, since such systems have to consider how human respond to speech warnings by changing their operating behavior under the influence of different warning characteristics.

Although this study was carefully prepared, there are still several limitations. First of all, the model was mainly validated with accident rates and response time since the published work only reported the accident rate as the objective index of warning response performance. Further work is needed to validate the detailed levels of the proposed model. Secondly, although the proposed mathematical model provides a promising tool to predict the effects of loudness of speech warnings on human performance, the influence of other acoustic properties, like frequency and pitch, and the threshold of intelligibility were not modeled. For example, the warning presented with a higher pitch (e.g., female voice) may have a different impact on human performance and subjective rating on warning urgency than that of a lower pitch (e.g., male voice). Meanwhile, there might be interactions between signal word choice and other acoustic factors. The current work assumed that the perceived urgency expressed by different signal words is relatively stable, but the perceived urgency might vary with the signal words presented at different pitch and frequency levels. To enhance the model in predicting speech warning acoustics and semantic properties on human behaviors, further work is needed to model the interaction among acoustic properties and the interaction between signal word choice and other acoustic properties. Furthermore, the current QN-MHP model did not account for individual differences, but may significantly contribute to the model application. For example, although the model predicted optimal lead time, loudness level and signal word for speech

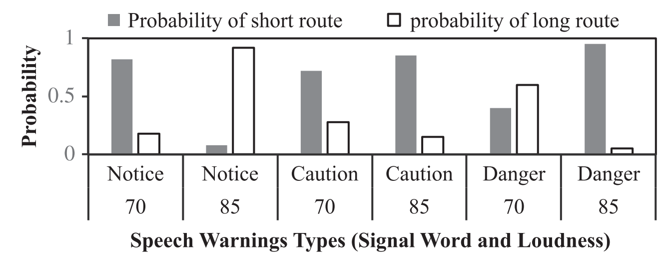


Fig. 10. The simulation results of route choice for warnings with different loudness level and signal words.

warnings, personality (e.g., aggressive vs. conservative drivers) may affect a driver's responses to speech warnings. Ideally, future model should consider individual differences and provide different system design suggestions according to individual characteristics instead of an average driver.

Appendix
The Q Online Learning Algorithm and Modeling of Learning Process

The Q online learning algorithm will be integrated with the QN-MHP to model the learning in route choice under the influence of warning loudness and word choice. The effect of speech warning parameters on reaction time ($I_{RT,i}$) and response error rate ($I_{E,i}$) is then modeled with the different route choices in the information processing. As it presented in the following equations (Equations 9.2 and 9.3,) [21], the choice of route is based on the updated Q value $Q^{t+1}_{(i,j)}$ in each transition:

$$Q^{t+1}_{T(i,j)} = Q^{t}_{T(i,j)} + \varepsilon\left\{r'_t + \gamma \max_k\left[Q^{t}_{T(i,k)}\right] - Q^{t}_{T(i,j)}\right\} [21] \tag{25}$$

$$Q^{t+1}_{E(i,j)} = Q^{t}_{E(i,j)} + \varepsilon\left\{r''_t + \gamma \max_k\left[Q^{t}_{E(i,k)}\right] - Q^{t}_{E(i,j)}\right\} [21] \tag{26}$$

where $Q^{t+1}_{(i,j)}$ is the online Q value if entity routes from server i to server j in $t+1$th transition. $\max_k[Q^{t}_{(i,k)}]$ denotes the maximum Q value routing from server j to next k servers ($k \leq 1$); r_t is the reward; γ is the discount parameter of routing to next server ($0 < \gamma < 1$). The time-saving reward (r'_t) is modeled as $r'_t = (1/w_q) + \mu_{j,t}$, where w_q is the waiting time in the queuing at the server; the error-saving reward r''_t is modeled as $r''_t = (1/(N_{\mathrm{error}(j,t)} + 1))$, where $N_{\mathrm{error}(j,t)}$ is the number of action errors of the previous entities made in the next server j at tth transition

$$N_{\mathrm{error}(j,t)} = N_{\mathrm{error}(j,t)} + 1 \times L/100 \times U_S.$$

Both $Q^{t+1}_{E(i,j)}$ and $Q^{t+1}_{T(i,j)}$ will contribute to the survival chance when human respond to warnings toward a potential hazard. Therefore, the choice of routes is determined by the sum of two Q values. Currently, it is assumed that Q value of the error-saving reward and the Q value of the time-saving reward has the same priority. If $Q^{t+1}_{E(i,j)} + Q^{t+1}_{T(i,j)} > Q^{t+1}_{E(i,k)} + Q^{t+1}_{T(i,k)}$, the entity will choose server j; if $Q^{t+1}_{E(i,j)} + Q^{t+1}_{T(i,j)} < Q^{t+1}_{E(i,k)} + Q^{t+1}_{T(i,k)}$, the entity will choose server k; and if $Q^{t+1}_{E(i,j)} + Q^{t+1}_{T(i,j)} = Q^{t+1}_{E(i,k)} + Q^{t+1}_{T(i,k)}$, the entity will choose the next server (j or k) randomly. The simulation results of probability of route choices is shown in Fig. 10.

References

[1] NHTSA, "2011 Traffic Safety Facts Data Summary Booklet," U.S. Dept. Transp., Washington, DC, USA, 2013. [Online]. Available: http://www-nrd.nhtsa.dot.gov/Pubs/811755DS.pdf.

[2] D.-Y. Cheng *et al.*, "The optimal sampling period of a fingerprint positioning algorithm for vehicle speed estimation," *Math. Prob. Eng.*, vol. 2013, 2013, Art. no. 306783.

[3] J. Santa, A. F. Gómez-Skarmeta, and M. Sánchez-Artigas, "Architecture and evaluation of a unified V2V and V2I communication system based on cellular networks," *Comput. Commun.*, vol. 31, no. 12, pp. 2850–2861, Jul. 2008.

[4] J. Edworthy and R. Hards, "Learning auditory warnings: The effects of sound type, verbal labelling and imagery on the identification of alarm sounds," *Int. J. Ind. Ergonom.*, vol. 24, no. 6, pp. 603–618, Oct. 1999.

[5] K. Momtahan, R. Hetu, and B. Tansley, "Audibility and identification of auditory alarms in the operating room and intensive care unit," *Ergonomics*, vol. 36, no. 10, pp. 1159–1176, Oct. 1993.

[6] R. D. Patterson and C. A. Authority, *Guidelines for Auditory Warning Systems on Civil Aircraft*. London, U.K.: Civil Aviation Authority, 1982.

[7] J. Serrano, L. L. Di Stasi, A. Megias, and A. Catena, "Effect of directional speech warnings on road hazard detection," *Traffic Injury Prev.*, vol. 12, no. 6, pp. 630–635, Dec. 2011.

[8] K. R. Laughery, "Safety communications: Warnings," *Appl. Ergonom.*, vol. 37, no. 4, pp. 467–478, Jul. 2006.

[9] R. G. Mortimer, "Effectiveness of a warning as measured by behavior change," *Proc. Hum. Factors Ergonom. Soc. Annu. Meet.*, vol. 51, no. 9, pp. 574–577, Oct. 2007.

[10] M. S. Wogalter, M. J. Kalsher, and B. M. Racicot, "Behavioral compliance with warnings: Effects of voice, context, and location," *Safety Sci.*, vol. 16, no. 5/6, pp. 637–654, Aug. 1993.

[11] M. S. Wogalter, M. J. Kalsher, and R. Rashid, "Effect of signal word and source attribution on judgments of warning credibility and compliance likelihood," *Int. J. Ind. Ergonom.*, vol. 24, no. 2, pp. 185–192, May 1999.

[12] J. Wan, C. Wu, and Y. Zhang, "The effects of lead time of collision warning messages on driver performance," *Proc. Hum. Factors Ergonom. Soc. Annu. Meet.*, vol. 58, no. 1, pp. 2088–2092, Sep. 2014.

[13] G. Abe and J. Richardson, "The effect of alarm timing on driver behaviour: An investigation of differences in driver trust and response to alarms according to alarm timing," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 7, no. 4/5, pp. 307–322, Jul.–Sep. 2004.

[14] J. D. Lee, D. V. McGehee, T. L. Brown, and M. L. Reyes, "Collision warning timing, driver distraction, and driver response to imminent rear-end collisions in a high-fidelity driving simulator," *Hum. Factors*, vol. 44, no. 2, pp. 314–334, Summer 2002.

[15] G. Abe and J. Richardson, "The influence of alarm timing on braking response and driver trust in low speed driving," *Safety Sci.*, vol. 43, no. 9, pp. 639–654, Nov. 2005.

[16] E. E. Wiese and J. D. Lee, "Auditory alerts for in-vehicle information systems: The effects of temporal conflict and sound parameters on driver attitudes and performance," *Ergonomics*, vol. 47, no. 9, pp. 965–986, Jul. 2004.

[17] J. Edworthy, E. Hellier, K. Walters, W. Clift-Mathews, and M. Crowther, "Acoustic, semantic and phonetic influences in spoken warning signal words," *Appl. Cogn. Psychol.*, vol. 17, no. 8, pp. 915–933, Dec. 2003.

[18] M. J. Traxler, *Introduction to Psycholinguistics: Understanding Language Science*. Chichester, U.K.: Wiley, 2011, pp. 105–116.

[19] J. L. McClelland and J. L. Elman, "Interactive processes in speech perception: The TRACE model," in *Parallel Distributed Processing*, vol. 2. Cambridge, MA, USA: MIT Press, 1986, pp. 58–121.

[20] C. Wu and Y. Liu, "Queuing network modeling of driver workload and performance," *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 3, pp. 528–537, Sep. 2007.

[21] C. Wu, M. Berman, and Y. Liu, "Optimization in brain?–Modeling human behavior and brain activation patterns with queuing network and reinforcement learning algorithms," in *Computational Neuroscience*. New York, NY, USA: Springer, 2010, pp. 157–179.

[22] C. Wu and Y. Liu, “Queuing network modeling of transcription typing,” *ACM Trans. Comput.-Hum. Interact. (TOCHI)*, vol. 15, no. 1, p. 6, May 2008.

[23] G. Zhao, C. Wu, and C. Qiao, “A mathematical model for the prediction of speeding with its validation,” *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 2, pp. 828–836, Jun. 2013.

[24] C. Wu and Y. Liu, “Queuing network modeling of the psychological refractory period (PRP),” *Psychol. Rev.*, vol. 115, no. 4, pp. 913–954, Oct. 2008.

[25] C.-J. Lin and C. Wu, “Mathematically modelling the effects of pacing, finger strategies and urgency on numerical typing performance with queuing network model human processor,” *Ergonomics*, vol. 55, no. 10, pp. 1180–1204, Oct. 2012.

[26] G. Zhao and C. Wu, “Mathematical modeling of driver speed control with individual differences,” *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 43, no. 5, pp. 1091–1104, Sep. 2013.

[27] C. Wu, Y. Liu, and C. M. Quinn-Walsh, “Queuing network modeling of a real-time psychophysiological index of mental workload—P300 in event-related potential (ERP),” *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 38, no. 5, pp. 1068–1084, Sep. 2008.

[28] Q. Ma, J. Jin, and L. Wang, “The neural process of hazard perception and evaluation for warning signal words: Evidence from event-related potentials,” *Neurosci. Lett.*, vol. 483, no. 3, pp. 206–210, Oct. 2010.

[29] J. Valls-Solé *et al.*, “Reaction time and acoustic startle in normal human subjects,” *Neurosci. Lett.*, vol. 195, no. 2, pp. 97–100, Aug. 1995.

[30] J. Qin and S. Han, “Neurocognitive mechanisms underlying identification of environmental risks,” *Neuropsychologia*, vol. 47, no. 2, pp. 397–405, Jan. 2009.

[31] F. Pulvermüller, M. Härle, and F. Hummel, “Walking or talking?: Behavioral and neurophysiological correlates of action verb processing,” *Brain Lang.*, vol. 78, no. 2, pp. 143–168, Aug. 2001.

[32] J. N. M. Fuster, “Network memory,” *Trends Neurosci.*, vol. 20, no. 10, pp. 451–459, Oct. 1997.

[33] T. D. Blumenthal, “The startle response to acoustic stimuli near startle threshold: Effects of stimulus rise and fall time, duration, and intensity,” *Psychophysiology*, vol. 25, no. 5, pp. 607–611, Sep. 1988.

[34] T. D. Blumenthal *et al.*, “Committee report: Guidelines for human startle eyeblink electromyographic studies,” *Psychophysiology*, vol. 42, no. 1, pp. 1–15, Jan. 2005.

[35] G. Martino and L. E. Marks, “Perceptual and linguistic interactions in speeded classification: Tests of the semantic coding hypothesis,” *Perception*, vol. 28, no. 7, pp. 903–924, 1999.

[36] C. L. Baldwin, “Verbal collision avoidance messages during simulated driving: Perceived urgency, alerting effectiveness and annoyance,” *Ergonomics*, vol. 54, no. 4, pp. 328–337, Apr. 2011.

[37] J. A. Foss, J. R. Ison, J. P. Torre, Jr., and S. Wansack, “The acoustic startle response and disruption of aiming: I. Effect of stimulus repetition, intensity, and intensity changes,” *Hum. Factors*, vol. 31, no. 3, pp. 307–318, Jun. 1989.

[38] E. Hellier, J. Edworthy, B. Weedon, K. Walters, and A. Adams, “The perceived urgency of speech warnings: Semantics versus acoustics,” *Hum. Factors*, vol. 44, no. 1, pp. 1–17, Spring 2002.

[39] P.-S. Jang, “Designing acoustic and non-acoustic parameters of synthesized speech warnings to control perceived urgency,” *Int. J. Ind. Ergonom.*, vol. 37, no. 3, pp. 213–223, Mar. 2007.

[40] I. S. Sigalovsky and J. R. Melcher, “Effects of sound level on fMRI activation in human brainstem, thalamic and cortical centers,” *Hearing Res.*, vol. 215, no. 1, pp. 67–76, May 2006.

[41] S. Uppenkamp *et al.*, “Locating the initial stages of speech—Sound processing in human temporal cortex,” *NeuroImage*, vol. 31, no. 3, pp. 1284–1296, Jul. 2006.

[42] K. B. Laughery, “Computer simulation of short-term memory: A component-decay model,” in *The Psychology of Learning and Motivation*, vol. 3, J. T. Spense and G. H. Bower, Eds. San Diego, CA, USA: Academic, 1970, pp. 135–200.

[43] V. Vorhold *et al.*, “The neural basis of risk ratings: Evidence from a functional magnetic resonance imaging (fMRI) study,” *Neuropsychologia*, vol. 45, no. 14, pp. 3242–3250, Nov. 2007.

[44] C. Gonzalez, B. A. Lewis, D. M. Roberts, S. M. Pratt, and C. L. Baldwin, “Perceived urgency and annoyance of auditory alerts in a driving context,” *Proc. Hum. Factors Ergonom. Soc. Annu. Meet.*, vol. 56, no. 1, pp. 1684–1687, Sep. 2012.

[45] J. Parmanen, “A-weighted sound pressure level as a loudness/annoyance indicator for environmental sounds—Could it be improved?” *Appl. Acoust.*, vol. 68, no. 1, pp. 58–70, Jan. 2007.

[46] C. D. Wickens, *Engineering Psychology and Human Performance*. New York, NY, USA: Harper Collins, 1992.

[47] C. Mulert, J. Gallinat, H. Dorn, W. M. Herrmann, and G. Winterer “The relationship between reaction time, error rate and anterior cingulate cortex activity,” *Int. J. Psychophysiol.*, vol. 47, no. 2, pp. 175–183, Feb. 2003.

[48] P. Keuss and M. Van der Molen, “Positive and negative effects of stimulus intensity in auditory reaction tasks: Further studies on immediate arousal,” *Acta Psychol.*, vol. 52, no. 1/2, pp. 61–72, Nov. 1982.

[49] C. L. Baldwin, “Cognitive implications of facilitating echoic persistence,” *Memory Cognition*, vol. 35, no. 4, pp. 774–780, Jun. 2007.

[50] A. W. Melton, “Implications of short-term memory for a general theory of memory,” *J. Verbal Learn. Verbal Behav.*, vol. 2, no. 1, pp. 1–21, 1963.

[51] J. Freedman and W. A. Rumbaugh, “Accuracy and speed of response to different voice types in a cockpit voice warning system,” Air Force Inst. Technol., Wright-Patterson AFB, OH, USA, 1983. [Online]. Available: http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html &identifier=ADA135595.

[52] H. Frenz and M. Lappe, “Absolute travel distance from optic flow,” *Vis. Res.*, vol. 45, no. 13, pp. 1679–1692, Jun. 2005.

[53] D. R. Montello, “The perception and cognition of environmental distance: Direct sources of information,” in *Spatial Information Theory a Theoretical Basis for GIS*. Berlin, Germany: Springer-Verlag, 1997, pp. 297–311.

[54] F. P. Redlick, M. Jenkin, and L. R. Harris, “Humans can use optic flow to estimate distance of travel,” *Vis. Res.*, vol. 41, no. 2, pp. 213–219, Jan. 2001.

[55] B. G. Witmer and P. B. Kline, “Judging perceived and traversed distance in virtual environments,” *Presence*, vol. 7, no. 2, pp. 144–167, Apr. 1998.

[56] B. R. Fajen, “Perceptual learning and the visual control of braking,” *Perception Psychophys.*, vol. 70, no. 6, pp. 1117–1129, Aug. 2008.

[57] D. Kadihasanoglu, R. D. Beer, and G. P. Bingham, “The dependence of braking strategies on optical variables in an evolved model of visually-guided braking,” in *From Animals to Animats 11*. Berlin, Germany: Springer-Verlag, 2010, pp. 555–564.

[58] C. L. Baldwin and J. F. May, “Loudness interacts with semantics in auditory warnings to impact rear-end collisions,” *Transp. Res. F, Traffic Psychol. Behav.*, vol. 14, no. 1, pp. 36–42, Jan. 2011.

[59] J. Hawkins, Jr. and S. Stevens, “The masking of pure tones and of speech by white noise,” *J. Acoust. Soc. Amer.*, vol. 22, no. 1, pp. 6–13, Jan. 1950.

[60] A. Weber and O. Scharenborg, “Models of spoken word recognition,” *Wiley Interdisc. Rev. Cogn. Sci.*, vol. 3, no. 3, pp. 387–401, May 2012.

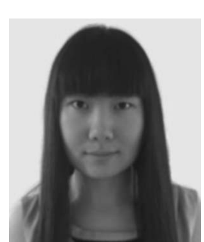
[61] M. G. Gaskell and W. D. Marslen-Wilson, “Integrating form and meaning: A distributed model of speech perception,” *Lang. Cogn. Process.*, vol. 12, no. 5–6, pp. 613–656, Oct. 1997.

[62] D. Norris, “Shortlist: A connectionist model of continuous speech recognition,” *Cognition*, vol. 52, no. 3, pp. 189–234, Sep. 1994.

[63] O. Scharenborg, “Modeling the use of durational information in human spoken-word recognition,” *J. Acoust. Soc. Amer.*, vol. 127, no. 6, pp. 3758–3770, Jun. 2010.

[64] P. A. Luce, “A computational analysis of uniqueness points in auditory word recognition,” *Perception Psychophys.*, vol. 39, no. 3, pp. 155–158, Mar. 1986.

[65] L. Ten Bosch, L. Boves, and M. Ernestus, “Towards an end-to-end computational model of speech comprehension: Simulating a lexical decision task,” in *Proc. 14th Annu. Conf. Int. Speech Commun. Assoc.*, 2013, pp. 2822–2826. [Online]. Available: http://pubman.mpdl.mpg.de/pubman/faces/viewItemOverviewPage.jsp?itemId=escidoc:1835410.



Yiqi Zhang received the B.S. degree in psychology from Zhejiang University, Hangzhou, China, in 2011. She is currently working toward the Ph.D. degree in industrial and system engineering with The State University of New York at Buffalo, Buffalo, NY, USA. Her current research interest includes studying driver behavior and performance in connected vehicle systems, the development of mathematical models of human response in the interaction with engineering systems, and addressing human cognition with their application in the design of intelligent transportation systems.

Changxu Wu (S'04–M'07) received the M.S. and Ph.D. degrees in industrial and operational engineering from University of Michigan, Ann Arbor, MI, USA, in 2004 and 2007, respectively. Since August 2007, he has been an Associate Professor with the Department of Industrial and Systems Engineering, The State University of New York at Buffalo, Buffalo, NY, USA, where he is also the Director of the Cognitive System Laboratory. He is interested in integrating cognitive science and engineering system design, particularly modeling human cognition system with its applications in system design, improving transportation safety, promoting human performance in human–computer interaction, and inventing innovative sustainable and smart energy systems with human in the loop. He has published more than 36 journal papers in these fields. He is an Associate Editor of IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATIONS SYSTEMS and *Behaviour & Information Technology*.

Jingyan Wan received the B.S. degree in automation and the M.S. degree in control science and engineering from Harbin Institute of Technology, Harbin, China, in 2007 and 2009, respectively. She is currently working toward the Ph.D. degree in industrial engineering with The State University of New York at Buffalo, Buffalo, NY, USA. Her research interests are driver behavior and performance in automated vehicles and ecodriving systems with human in the loop.

AUTHOR QUERY

NO QUERY.