


A Research of Speech Signal of Fire Information Display Interface

View metadata, citation and similar papers at core.ac.uk

brought to you by  CORE

provided by Institutional Repository of Institute of Psychology, CAS

Liang Zhang , Xianghong Sun , and Thomas Plocher

¹ State Key Laboratory of Brain and Cognitive Science, Institute of Psychology,
Chinese Academy of Sciences,
Beijing 100101, China

² Honeywell ACS Labs, Minneapolis MN 55418, USA
zhangl@psych.ac.cn, tom.plocher@honeywell.com

Abstract. This study was conducted to investigate the effect of speech rate and tune on intelligibility of fire information words and sentences under the conditions with different levels of noise. The result showed that the types of signals and noise levels affect the intelligibility significantly. The appropriate tune for fire information display interface is mezzo-soprano. The appropriate voice rate is 5 characters per second for words display, 7 characters per second for usual sentences display and 6 characters per second for the sentences with numbers display.

Keywords: fire information display interface; speech signal; tune; speech rate; noise occultation.

1 Introduction

Fire alarm system is an essential part of high buildings in modern times, which helps firefighters' detection more efficiently and reduces the casualty [1]. The fire information display of the alarm system can show the information inside the building, e.g. build structure, fireplug locations, firing time and locations, etc.

The firefighters have to make decisions depending greatly on huge amount of information provided and reach to their decisions in limited time. It was seen that using the single visual display is likely to result in users' visual overloading, inefficiency of information processing and even mistakes of decision making.

Auditory system is another important sensory system for getting information, which is the major complement to visual system. Moreover, human responds to auditory stimulus is faster than visual stimulus [2]. So using speech signals could be promising to solve the problems mentioned above.

The issue of using speech signals has come to engineering psychologists' attention [6,7]. Intelligibility is the precondition of signal application and the essential evaluating index [3]. The speech rate and tune are the essential properties of speech signals. At present, there have been some related researches abroad [8] but rarely in China. Unfortunately, there are also diversity results among the different language

using. And Chinese pronunciation, grammar, expression etc. are all unique from other languages. So the parameters of Chinese speech signals should be studied focused on the property of Chinese.

In addition, the types of phrases, such as the acquaintance to the content, the length of the speech, whether include numbers, etc., also affect the signal intelligibility. There are several related researches (Liu Baoshan, 1995; Zhang Tong, 1997) in China, but all were aviation warning system voice. The signals are verb phrases contained 2-5 characters, without numbers, which are highly professional [4]. Considering the different experimental materials and background, the results, i.e. the speech parameters are likely different. So far, there is not any human factors standardization of speech signals. That's the meaning of this study.

Another important factor considered in this study is the background noise, which has been evidence to affect on auditory signals [5]. The speech signal parameters should vary according to its using situation. Considering the high noisy background in fire emergency, we concerned the noise effect on speech signal, and the appropriate parameter under the high level of noise.

To sum up, this study aims to investigate the speech rate and tune under different levels of noise background, using the fire information phrases. And the result is expected to be helpful for the design of fire information display interface.

2 Method

Design. A 4-factor, within-subjects design was conducted. Each participant experienced a total of three experimental sessions. The factors and their levels are as follows.

Table 1. Factors and their levels

| Factor | A | B | C | D |
|---------|-------------------------------|---------------|------------------|---------------------|
| | Speech Rate (characters/s) | Tune | Speech type | Noise Level (dB) |
| Level 1 | 5 | Soprano | Noun | 0 |
| Level 2 | 6 | Mezzo-soprano | Numeral | 60 |
| Level 3 | 7 | Tenor | General sentence | 80 |
| Level 4 | 8 | Baritone | Numeral sentence | |
| Level 5 | 9 | | | |

Materials. 80 sentences or words, derived from the firefighters' interview and Chinese Fire Fighting Constitutions, are balanced distributed to the treatments combined of A, B, C factors. The sentences' length is 2-4characters, while the words' 6-10 characters.

Index of speech and noise. The speech was monotone without emotion, recorded by t male and female announcers in an isolated booth. The experimental phrases were processed from the records by the sound software, and its time accuracy is +1ms.

The noise background was generated from pink noise. The SNR (Signal Noise Ratio) was 5 dB. The dB size was measured by a SL-4001 sound instrument, the error size of which is +1dB, accurate to 0.1dB.

Participants. The participants in this experiment were 36 males, aged from 20-27 years old. All participants reported well audition and fluent mandarin.

Procedure. There was three experimental sessions, and each session under one noise level. In every session, participants can hear some sentences (or words) from the earphone. There was a pause after each sentence (or word), the participants were asked to repeat loudly what they just heard as far as they can. There were 5 minutes break between two sessions. And the sessions sequence were Latin square balanced.

Data record and process. The participants’ respond accuracy and reaction time were collected, and then processed by spss.11.0.

3 Results

There are 34 data available. In the analysis of accuracy, the incorrect, incomplete or vacant reactions were recorded as “0”, while the correct reactions were recorded as “1”. In the analysis of reaction time, the “0” reaction time was filled with the mean value of that treatment combination.

3.1 Effect of Noise

A repeated-measures statistical analysis was conducted and found the difference of accuracy ($F (2 , 66) =14.04 , p<0.001$ and the RT ($F (2 , 66) =14.04 , p<0.001$) are both significant.

The accuracy of 0 dB is significant higher than the other two noise levels. But there is no difference between the 60 and 80 dB. The mean of accuracy is showed in figure1 (the left).

The multiple comparisons on RT showed the differences between any two noise levels. The 80dB is the shortest, then 0 dB, and 60dB is the longest. See figure1 (the right)

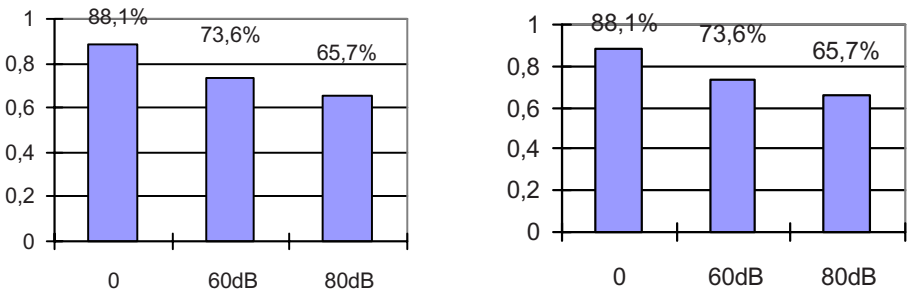


Fig. 1. Means of accuracy (the left) and RT (the right) (ms) under different levels of noise

3.2 Speech Rate

The accuracy descended with the speech rate ascended. Furthermore, the descent velocity varies under different noise. There is a interaction between speech rate and noise level, $F_{\text{accuracy}}(8,264)=6.16$, $p<.001$; $F_{\text{RT}}(8, 264)=29.07$, $p<.001$. The accuracy descends the most rapidly under the 80dB noise.

Considering the high noisy background in fire emergency, we analysis the data under the high noisy level father more.

Under 80dB, the accuracy of 5 levels of speech rate descend linearly, $F(4, 132)=64.55$, $p<.001$. The difference between any two levels is significant. See figure 2(the left).

The RT varied also, but not linearly, $F(4, 132) = 10.975$, $p<.001$. The multiple comparisons showed that the RT was short when 5, 6, 7 characters per second, and there is no significant difference among the three levels. But when the rate quickened up to unclear (the accuracy below 60%), the RT increased sharply. The RT on 8,9 characters per second were longer ($p<.01$). See figure 2(the right).

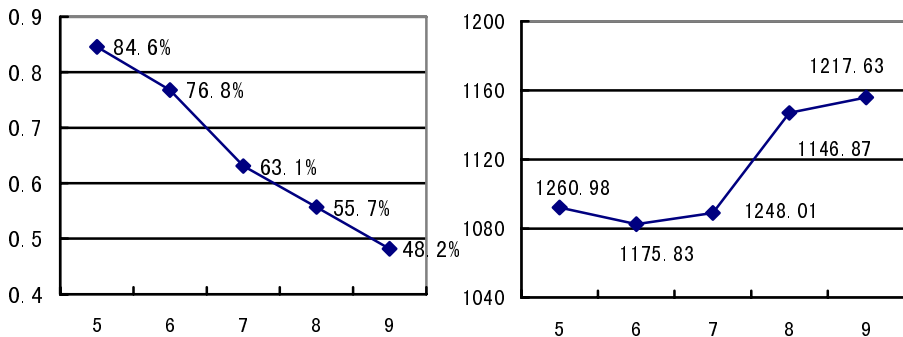


Fig. 2. Means of accuracy (the left) and RT (the right) (ms) on each speech rate (characters/s)

3.3 Tune

The male speech is more likely to be occulted than female's. Moreover, there is an interaction between tune and noise. The difference among tunes is greater under the noisy background and less different when there's no noise. Mezzo-soprano's accuracy is much higher than the others, especially under the 80dB noise. The tune major effect under 80dB is significant, $F_{\text{accuracy}}(3,99)=40.245$, $p<.001$; $F_{\text{RT}}(3,99)=6.711$, $p<.001$.

The RT of each tune under the 80dB noise level is showed as Table 2. Mezzo-soprano's accuracy is the highest, and tenor and mezzo-soprano's RT are shorter, but there is no statistical difference between them.

Generally, the mezzo-soprano is the optimal for speech signal.

Table 2. RT (ms) of each tune under the 80dB noise level

| Tune | Soprano | Mezzo-soprano | Tenor | Baritone |
|--------|---------------|---------------|---------------|----------------|
| RT(ms) | 1138.98±11.35 | 1101.80±9.78 | 1093.98±11.35 | 1118.400±10.41 |

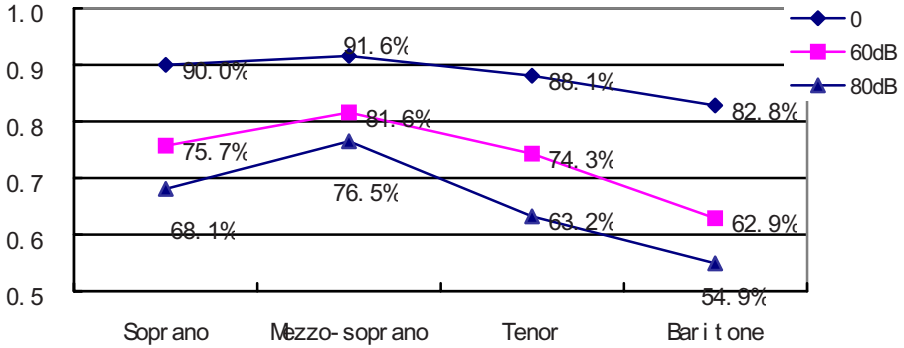


Fig. 3. Accuracy of each tune under the three levels of noise

3.4 Speech Type

The main effect of speech type is significant, $F_{accuracy}(3, 99) = 4.504, p < .01$; $F_{RT}(3, 99) = 226.33$. And the interaction between type and speech rate is significant, $F_{accuracy}(12, 396) = 17.391, p < .001$; $F_{RT}(12, 396) = 51.37, p < .001$. The four types' optimal speech rates are somewhat different.

Farther analysis showed that: (a) looking at the aspect of speech length, the sentences' accuracy descend faster than words', while their RT ascending is similar; (b) looking at the aspect of the content, signals containing numbers descend faster than those without numbers, and the RT ascending faster also. Considering the factors above, the four types' appropriate speech rate are as follows:

Table 3. The optimal rate(characters/s) of each speech signal type

| Type of speech signal | Noun | Numeral | General Sentence | Numeral Sentence |
|----------------------------|------|---------|------------------|------------------|
| Speech rate (characters/s) | 5 | 5 | 7 | 6 |

4 Discussion

4.1 Effect of Noise

In the researches of display design, it's very important to conform the experimental condition to the use situation. Therefore, the design of speech signal on fire display shouldn't ignore the high noisy background.

It's obvious that the noise occultation affect the speech signals intelligibility. The intelligibility descended with the noise heightened. That means the speech rate should be relatively slowed in the noisy environment.

Further more, the noise occultation is close related to the voice properties, such as pitch or audio-frequency etc. According to the previous research, basso is prone to be occulted. That's the reason why we didn't make the bass or alto as an experimental level. The results also showed female signals are superior to males', and the superiority comes greater in noisy environment.

Besides, we compared the high noisy level with the low noisy and without noise level. The different results among the three conditions may indicate a general trend of noise effect to speech signal, though the details still need more researches.

4.2 Acquaintance Effect to Intelligibility

Contrastively, the accuracy in current study is inferior to some researches before. That can be due to the unfamiliarity to experimental materials. The experimental phrases are highly professional in the previous researches, so participants were asked to learn the phrases beforehand. But the specialization of this experimental phrase is much lower. Thus we didn't request participants to learn the phrases. That's the major reason that leads to the inferior intelligibility.

Reviewing two related researches, we found the acquaintance effect also. Comparing Liu Baoshan's (1995) study with Zhang Tong's (1997), the former accuracy was higher than the latter, though their materials were both fighter warning system voice. It is mostly because the former participants were pilots while the latter were undergraduates.

Moreover, we found the practice effect in our research. It suggested the acquaintance to signals could affect their intelligibility. So the further study plans to use firefighters as participants. And the better intelligibility is anticipated because the firefighters are much familiar with the fire phrases. The relationship between acquaintance and signal intelligibility can be research farther by comparing the current results with the firefighters' research.

4.3 Contrast

Table 4 lists the contrast among the results in current study, the criterion of American military systems design and the voice parameters in Chinese fighter warning system.

As mentioned above, there isn't a uniform criterion for speech signal design in China. The speech signal not only varies with the languages, but also varies with the speech properties (e.g. content, length, and receiver's attention state etc.).

Compare with the warning voice, the speech signal on display interface is more complex. It can express more information, so its content and length are more various. Moreover, there is a distinction between them. That is the receiver's attentive state. The warning voice often broadcast without any preparation. So it requests to attract receivers' attention rapidly and forcefully. On the contrary, receivers usual concerned with the display when they use it, so they would be easier to receive the speech signal information. That's why the signal speech rate is faster than the warning voice.

Table 4. Contrast with other parameters of auditory signal

| | Type of Task | Tune | Voice rate | RT(s) |
|--|--------------|---------------|-------------------|-------|
| American voice warning | | Female | 156~178 words/min | |
| Chinese voice warning In fighter cockpit | Single-task | Mezzo-soprano | 180~300 ch/min | |
| | Double-task | | 240~360 ch/min | 2.181 |
| Speech signal in fire interface | Single-task | Mezzo-soprano | 300~420 ch/min | 1.102 |

4.4 Further Research

This study simply concentrated on the auditory signal design. In fact, multiple display interfaces is more efficient than single visual or auditory interface. Therefore, further researches plan to explore the combination of visual and auditory interface. The multiple interfaces will be more efficient and humanized than single interface, and its application is anticipant to be exciting.

5 Conclusion

The result showed that the types of signals and noise levels affect the intelligibility significantly. Under the high noisy background, the optimal tune for fire information display interface is mezzo-soprano. The optima speech rate is 5 characters per second for words display, 7 characters per second for usual sentences display and 6 characters per second for the sentences with numbers display.

References

1. Zhijian, F.: Development of human-machine interaction:multimedia and multisensory. *Human Factors* 2, 34–38 (1998)
2. Quan, P.: Design of application Interface based on the human cognition. *Computer engineer and application* 19, 148–150 (2001)
3. Tong, Z.: The research of voice rate for warning system. *Chinese Journal of Applied Psychology* 1, 34–39 (1997)
4. Xioxia, G.: Selection of Chinese phrase for fighter voice warning system. *Chinese Journal of Aviation Medicine* 5, 230–231 (1994)
5. Baoshan, L.: Determination of design parameters of synthesized Chinese voice speech for auditory warning system in fighter cockpit. *Human Factors* 1, 24–27 (1995)
6. Minnesota, S.J., California, W.A.: Critical issues in the development of speech technology for manmachine
7. Integration. In: *Proceedings of 26th Annual Meeting of the Human Factor Society* (1982)
8. Steeneken, H.J.M., Langhout, G.: Voice interactive system in severe noise condition, (1985) ADP005603
9. Simpson, C.A.: Synthesized speech rate and pitch effects on intelligibility of warning messages for pilots. *Human Factor* 5, 32–33 (1984)
10. MIL-STD-1472D. Human engineering design criteria for military systems. *Equipment and facilities*, pp. 39–46 (1989)