Chinese speaking times
Phoebe S.K. Lee, Alan H.S. Chan*

Department of Manufacturing Engineering and Engineering Management, City University of Hong Kong, Tat Chee Avenue, Kowloon Tong, Hong Kong

Received 13 March 2001; received in revised form 19 August 2002; accepted 14 October 2002

Abstract

This paper presents the results of elemental standard Chinese speaking times per character and explores the factors that may affect these results. Over a million characters were recorded in the study. The collected data was analyzed with analysis of variance and the results showed that average speaking times for Cantonese and Mandarin are different. It is about 231 ms/character for Cantonese and 224 ms/character for Mandarin. An interesting dimension of this research is the interaction of language with factors of time session, format, gender and news category. This study indicates that these factors needed to be considered during compilation of standard speaking time for use. The result of this study is useful in generating a new and important category of communication basic times for standardizing and evaluating speech performance of workers in this competitive century.

Relevance to industry

The establishment of standard Chinese speaking times will have a significant contribution to the work design and methods engineering in the human communication process which forms a basic component of vocalization tasks in a number of service industries, finance industries, and business services.

© 2002 Elsevier Science B.V. All rights reserved.

Keywords: Work design; Standard speaking time; Cantonese; Mandarin; Predetermined time systems

1. Introduction

Predetermined time systems (PTS) are standard data systems with tabulated values of basic times for use in a wide variety of operation and process planning applications. The PTS has been an important technique in the field of methods and work measurement since the pioneering work of Frederick Winslow Taylor in 1881. Other than the time study technique, the PTS has a wide range of acceptance in industries because of its practicality and convenience of estimating task times required for an operation before the operation takes place. Approximately 50 different PTS have been devised in the past to provide established synthetic values (Niebel and Freivalds, 1999; Aft, 2000). Though these systems were developed with different usages and considerations in mind, they all share the common aim of establishing labor standards for basic motions for the evaluation and measurement

*Corresponding author. Tel.: +852-788-8439; fax: +852-788-8423.
E-mail address: alan.chan@cityu.edu.hk (A.H.S. Chan).
of productivity (Barnes, 1980; Whitmore, 1987; Sellie and Worrall, 1992; Mundel and Danner, 1994; Meyers, 1999; Niebel and Freivalds, 1999).

Time standards are known to be key elements of management information systems for process planning and improvement (Wrennall, 1999). However, most of the basic times available in the prevalent PTS are targeted for the measurement of physical activities and movements exercised by the fingers and hands of human operators. Other than the basic motion times, the principles of motion economy and fundamental motion analysis techniques have also proved successful in work design assignments in many industrial applications. There are different kinds of PTS for applications ranging from highly repetitive short-cycle operations to less repetitive long-cycle activities. A quick comparison of a few traditional PTS like work-factor and methods-time-measurement (MTM) showed that other than the system specific elements, there are also many physical motions like “move”, “get”, and “put” commonly included in these systems.

As the nature of human tasks has shifted from emphasis on physical activity to emphasizing cognitive skills, the business environment has changed and research in work design needs to be brought up to date. The development of time standards for processes that are primarily mental, like speaking, reading, and writing, has become necessary. For mental processes, the Mento-Factor System has been developed for establishing labor standards for tasks with major mental content like proofreading, visual inspection, problem solving, and calculations. Another PTS, MODular arrangement of predetermined time standards (MODAPTS) also contains basic times for fundamental mental processes as well as physical motions. It is unique in that it includes a new element “vocalize (V)” related to human speech, which refers to vocalizing or speaking an English word (Heyde, 1983; MODAPTS, 2000).

Speech is a system of audible and visible signs utilizing many muscles and tissues for communicating ideas amongst humans. It is also a form of human behavior that utilizes physical, psychological, semantic, and linguistic factors. In industry, differences in speech rates from different vocalizers and in different work situations are easily noticed. We have all experienced speech that is either too fast or too slow, and that causes annoyance, discomfort, or even failure to receive important information. Such speech is undesirable and defeats the purpose of speech communication, with the possible consequence of losing time and money. Although the basic times for vocalizing English are available and commonly used in industries such as banking, insurance, government authorities, and manufacturing, no similar system of data has been developed for use in the Chinese-speaking environment. The basic time unit for use in MODAPTS is MOD, which equals 0.129 s. The method study analysts use “V3” to represent that the average time to vocalize a word is 3 MODs (0.387 s). If the average time to speak a word for taxi dispatchers is 0.32 s, then a more realistic figure of V2.5, instead of V3, needs to be used. Because of the phonetic and formational differences between the English and Chinese language systems, the existing PTS for speaking of English (MODAPTS) cannot provide meaningful values for use in communication tasks involving Chinese characters. There is then an obvious and urgent need for setting up similar speaking time standards for the Chinese working environment.

In 1955 the government of the People’s Republic of China proclaimed a national language called Putonghua. It is a combination of Beijing dialect, the grammar of Northern Mandarin, and the vocabulary of modern vernacular literature. Based on the sound structures, there are seven major dialect groups in China, viz. Mandarin, Wu, Xiang, Gan, Hakka, Min, and Yue (Li and Thompson, 1981). The population of China is about 1275 million (21% of the world population) (United Nations Population Information Division, 2001). There are about 844 million native speakers of Mandarin and altogether 975 million people worldwide (16%) speak Mandarin (Grimes, 1996), compared with the second major group of 478 million (8%) English speakers. Amongst the seven major Chinese dialect groups, Yue (Cantonese) has a very special characteristic in that it has the most difficult tonal system. Cantonese is widely used in the southern part of China and Hong Kong, at which a high level of business and industrial activities are taking place (Taylor and
Taylor, 1995; Pennington, 1998). The number of Cantonese speakers is approximately 70 million in the world (Grimes, 1996).

It is known that the Chinese culture has a peculiar and enduring nature, which differs in many fundamental ways from Western culture. One of the most obvious differences is that Chinese language is non-agglutinative and it is a tonal and monosyllabic system (Tseng, 1993). There are 32 fundamental types of strokes in Chinese writing (Stallings, 1976; Gu, 1994). A stroke is strictly a building block of the shape of a Chinese character; it neither codes any sound nor represents any meaning. Several strokes form a character. A character is known as hánzi (“Han/Chinese graphs”) and it has meaning. All the characters are monosyllabic. One or several characters form a word. Different combinations of characters have different meanings, giving an unlimited number of words, which in turn combine to give an unlimited number of sentences (Lee, 1997). Regarding pronunciation, each Chinese character consists of three components: consonant, vowel and tone. For example, the character 他 (mother) is represented as mā in Mandarin and mā’ in Cantonese. Although there are differences found even within the different groups of Chinese dialects, their basic forms are similar and the phonemes of the seven major dialects are basically the same (DeFrancis, 1990). A Chinese character is a logograph. It represents a morpheme, and hence there should be as many characters as there are morphemes in Chinese. It was reported that there are about 48,000 different characters in the Chinese dictionary “Zhōnghuá dà zǎidiǎn” (Norman, 1988). However, only a few thousand are commonly used in daily life. In one count made in Taiwan, the number of characters in daily use is 4532 (Liu et al., 1975) while there are only 3000 characters commonly used in Hong Kong (Zhang, 1968). There have only been a few studies on the investigation of Chinese speaking. For instance, Goonetilleke et al. (1999) and Hoosain and Salili (1987) studied the memory span with Chinese speakers; Luximon et al. (2001) and Fu et al. (1998) analyzed the performance of Chinese speech recognition; and some other researchers (Chang, 1998; Chen, 1999; Zhou and Marslen, 2000) worked in the field of Chinese psycholinguistics; however none of them was done in the context of work science and work design.

Speech communication forms a basic component of jobs in a number of service industries, finance, and every aspect of business. For example, bank staff need to talk to customers at service counters or on the telephones to handle financial transactions. Customer service personnel in a domestic appliance maintenance company must converse with customers seeking technical services. By means of speech, insurance agents explain policy details and answer the inquiries of customers. For all these and similar tasks, the establishment of speaking time standards is essential for the estimation of the call handling capacity for telebanking, assessment of human effectiveness, and manpower planning. By means of speech, newscasters of television and radio broadcasting channels deliver news to the audience. Given the large Chinese speaking population and the changing needs of human skills for industries, the present study was designed and conducted with the aim of establishing basic speaking time estimates for Cantonese and Mandarin, and the examination of the possible effects that certain factors may have on these time values. The results of this study will provide a useful databank for use in work measurement, which is fundamental to the study of work science, and will contribute to the assessment and improvement of business practices.

2. Methodology

It was estimated that a large proportion of human-generated information is made by speech and that much of this is in the form of television and radio broadcasts (Morrison and Morrison, 1998). As newscasters are usually professional and articulate speakers who can utter clear and easily perceivable syllables and words, their speeches were recorded for analysis of human speech communication (Wang, 1997; Renals et al., 2000). Extensive experiments on broadcast news speech collected in Taiwan were also reported (Wang, 2000). In this study, the speeches of newscasters on seven television and radio
broadcasting channels in Hong Kong were recorded. It was believed that the choice of newscasters satisfied the requirement of selecting a “qualified employee” for establishing a fair and accurate time standard in which the term “qualified employee” is defined as “a representative average who is fully trained and able satisfactorily to perform any and all phases of the work involved, in accordance with the requirements of the job under consideration” (Niebel and Freivalds, 1999). Also newscasters are able to perform speech communication tasks to the general public at a normal rate which is defined as “the effective rate of performance of a conscientious, self-paced employee when working neither fast nor slow and giving due consideration to the physical, mental, or visual requirements of the specific job.”

In addition to the main objective of establishing basic speaking time estimates for Cantonese and Mandarin, the study was also designed with the objectives of examining time difference for other logically linked and naturally associated factors like audio and video format, time of the day, gender, and news category. There were hence two languages (Cantonese and Mandarin), four time sessions (morning, afternoon, evening, and late evening), two formats (video and audio), two genders, and seven news categories to be investigated. The possible influence of these factors on speech time was analyzed accordingly with an analysis of variance (ANOVA) considering the main factors of time session (S), format (F), language (L), gender (G) and news category (N) (Table 1). Other than the gender and news category, all the main factors were found to be significant. Regarding the session factor, the speaking time for the afternoon (214 ms) was the shortest while that of the late evening (254 ms) was the longest. The difference might be due to the fact that there is not much tension and stress in the late evening and the pace of life is generally not so hectic at that time session (Levine, 1997). In the daytime humans are always occupied with frenzied and hectic schedules and it is in the evening and late evening that they can have a chance to take a break from the day’s routines. The average speaking time for the evening session, 231 ms was close to the overall mean 230 ms. Duncan’s multiple-range test showed that each of the four session means was significantly different from the others (p < 0.05).

3. Results and analysis

Altogether 11,878 sentences (1,422,278 characters) over a time duration of 89 hours 16 minutes 30 seconds were collected. The ratio of speaking time and number of characters for a sentence gave the average speaking time per character. For removing any errors arising during recording or transcribing, the raw data set was subject to a 3σ limit control. After the scrutinizing process, 11,771 data sets were found lying within the upper (586 ms) and lower control limit (0 ms). The scrutinized data set follow a normal distribution (Kolmogorov–Smirnov test, p = 0.05) with an overall mean and standard deviation of 230 and 59 ms, respectively. For further analysis, the news was classified into seven categories, viz., political, economics, crimes, health, social welfare, sports, and others. The analysis of average speaking time was started with Analysis of Variance (ANOVA) considering the main factors of time session (S), format (F), language (L), gender (G) and news category (N) (Table 1).
need only to announce the news. However, in the video format, they need also to watch a number of cameras and pay attention to the cues provided by the workers around, which would then slow down the speaking rate. For the language, the speaking time of Mandarin (224ms) was shorter than that of Cantonese (231ms). It was also noted in the ANOVA that all the two-way interactions were significant. The entire two-way interaction plots for the five factors are shown in Fig. 1. Other than the two-way interactions, significant three-way interactions of session × format × gender (p<0.001), session × language × gender (p<0.01), and session × gender × news category (p<0.005) were significant. However, for the sake of practicability and convenience during data application, it is recommended to consider only two-way interactions during standard time compilation. As the language factor interacted significantly with the other four factors, the basic Chinese speaking times at various levels of these factors were so derived for practical use (Table 2).

4. Discussion and conclusion

The basic times in Predetermined Time Systems can be used in many industrial applications. Examples include the synthesizing of operation time standard in operation planning and the
determination of manpower planning. The established time standards provide criteria for operation and skill evaluation, as well as a training tool for vocalization and production of intelligible speeches in customer service industry. Given the change of task demands from manual and physical
effort to cognitive and mental resources for human workers, there is certainly a strong need of knowing the basic times for the human communication processes with the use of language. The speaking time standard is hence particularly useful and important for work standardization and job design in business nowadays. Similar to other PTS, an effective use of the speaking times generated in this study relies on a faithful adherence of the procedure that analyzes the operation or method into the basic elements. Each element is then assigned a predetermined time standard dependent on the nature of the element and the conditions under which the element is performed. For example, the application of element “Move” in MTM needs to consider four factors for standard time establishment, viz., distance of move, case of move, type of move, and weight to move (Karger and Bayha, 1987). Similarly, in establishment of standard speaking time, the factors of language, session, format, gender, and news category need also be considered for data compilation.

It is not surprising that the language factor was significant here, as Cantonese and Mandarin are really different spoken languages although they came from the same origin. The results showed that they should be separately handled during the establishment of standard speaking time. Mandarin speaking is found to be 3% faster than Cantonese. If basic time is extracted without consideration of other specific factors, values of 231 ms for Cantonese and 224 ms for Mandarin can be used. However, other than the language factor, speaking time also varied with the session and format.
of their speech should be of high standard and they are conscientious, self-paced employees working neither fast nor slow (Niebel and Freivalds, 1999). In many industries, differences of speech rate are easily noticed from different vocalizers and in different work situations causing the receivers annoyance, discomfort, or even failure to receive important information. This is totally undesirable and the purpose of speech in the communication process is defeated. Also we should be cautious in applying the results collected in this study to different work situations or subject groups. For example, the speaking rates established here are not necessarily applicable to a hotline for elderly as it was shown that the age of listeners also affects speech reception and intelligibility, particularly for the people above 60 years of age (Bergman et al., 1976). Also, there may be some other factors like environmental noise that may affect the optimum speaking rate. All these prompted the need of further research for improving the usability and reliability of data for optimum speaking speed in future. Apart from the standard Chinese speaking time, there is also a need to establish similar standards for other communication processes like writing and reading in the Chinese environment.

Acknowledgements

The work described in this paper was supported by a grant from City University of Hong Kong (Project No. 7001367).

We would like to thank Dr Michael Hui for providing useful advice on the data acquisition.

References


Table 2
Basic speaking times (ms/character) at different levels of session, format, gender, and news category factors

<table>
<thead>
<tr>
<th>Factor</th>
<th>Level</th>
<th>Language</th>
<th>Cantonese</th>
<th>Mandarin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Session</td>
<td>Morning</td>
<td>225</td>
<td>227</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Afternoon</td>
<td>210</td>
<td>224</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Evening</td>
<td>231</td>
<td>222</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Late evening</td>
<td>264</td>
<td>183</td>
<td></td>
</tr>
<tr>
<td>Format</td>
<td>Audio</td>
<td>228</td>
<td>245</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Video</td>
<td>237</td>
<td>183</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>Male</td>
<td>231</td>
<td>216</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>234</td>
<td>212</td>
<td></td>
</tr>
<tr>
<td>News category</td>
<td>Political</td>
<td>229</td>
<td>214</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Economics</td>
<td>233</td>
<td>214</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Crimes</td>
<td>228</td>
<td>222</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Health</td>
<td>247</td>
<td>204</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Social welfare</td>
<td>236</td>
<td>199</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sports</td>
<td>225</td>
<td>226</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Others</td>
<td>229</td>
<td>218</td>
<td></td>
</tr>
</tbody>
</table>


Pennington, M.C., 1998. Language in Hong Kong at century’s End. Hong Kong University Press, Hong Kong.


