Speech Adjustments for Room Acoustics and Their Effects on Vocal Effort

Pasquale Bottalico, *East Lansing, Michigan*

Summary: Objectives. The aims of the present study are (1) to analyze the effects of the acoustical environment and the voice style on time dose (D_{t_p}) and fundamental frequency (mean f_0 and standard deviation std_f_0) while taking into account the effect of short-term vocal fatigue and (2) to predict the self-reported vocal effort from the voice acoustical parameters.

Methods. Ten male and ten female subjects were recorded while reading a text in normal and loud styles, in three rooms—anechoic, semi-reverberant, and reverberant—with and without acrylic glass panels 0.5 m from the mouth, which increased external auditory feedback. Subjects quantified how much effort was required to speak in each condition on a visual analogue scale after each task.

Results. (Aim1) In the loud style, D_{t_p}, f_0, and std_f_0 increased. The D_{t_p} was higher in the reverberant room compared to the other two rooms. Both genders tended to increase f_0 in less reverberant environments, whereas a more monotonous speech was produced in rooms with greater reverberation. All three voice parameters increased with short-term vocal fatigue. (Aim2) A model of the vocal effort to acoustic vocal parameters is proposed. The sound pressure level contributed to 66% of the variance explained by the model, followed by the f_0 (30%) and the modulation in amplitude (4%).

Conclusions. The results provide insight into how voice acoustical parameters can predict vocal effort. In particular, it increased when SPL and f_0 increased and when the amplitude voice modulation decreased.

Key Words: Voice acoustical parameters–Room acoustics–Vocal effort–Vocal fatigue–Speech adjustments.

INTRODUCTION

Although speech acoustic parameters are strongly related to physiological factors such as vocal tract size, vocal fold length, and lung capacity, speakers can adjust their voice to achieve the desired vocal output. This vocal output is affected by various factors such as the type of environment[1,2] and interlocutor.[3]

Fundamental frequency mean (f_0) and standard deviation (std_f_0) appear to be affected by the room acoustics and in particular by the reverberation time (T_{30}).[4] This parameter is the duration required for the space-averaged sound energy density in an enclosure to decrease by 60 dB after the source emission has stopped.[4] The effect of the environment on speech acoustics was investigated by Pelegrín-García et al,[1] who considered the talker-listener distance. Thirteen male talkers were recorded in four different environments: an anechoic chamber, a lecture hall, a corridor, and a reverberant room with reverberation times averaged between 500 Hz and 1000 Hz ($T_{30,\ 0.5–1\ kHz}$) of 0.04 second, 1.88 seconds, 2.34 seconds, and 5.38 seconds, respectively. The parameters analyzed by the authors included phonation time ratio, which is the ratio between the phonation time (total duration of voiced frames) and the running speech time (total duration of the recording without pauses longer than 200 ms), and f_0 mean and standard deviation. The phonation time ratio changed significantly among rooms. In the anechoic room and the reverberant room, it was higher of about 10% compared to the rooms in the lecture hall and the corridor. The f_0 mean and standard deviation decreased with an increase in the reverberation time.

Accepted for publication October 4, 2016.

From the Voice Biomechanics and Acoustics Laboratory, Department of Communicative Sciences and Disorders, Michigan State University, East Lansing, Michigan.

Address correspondence and reprint requests to Pasquale Bottalico, Voice Biomechanics and Acoustics Laboratory, Department of Communicative Sciences and Disorders, Michigan State University, East Lansing, MI. E-mail: pb@msu.edu

Journal of Voice, Vol. 31, No. 3, pp. 392.e1–392.e12

0892-1997

http://dx.doi.org/10.1016/j.jvoice.2016.10.001

Phonation time (D_{t_p}) appears to increase under more reverberant conditions, with a consequent increase in vocal fatigue.[2] The influence of different acoustic environments on the duration of voicing and silence frames in continuous speech was investigated by Astolfi et al.[2] Part of their study involved the analysis of phonation time in percent (D_{t_p}) from free speech of 5 minutes in duration, which was performed by 22 university students in a reverberant room and a semi-anechoic room ($T_{30,\ 0.5–2\ kHz}$ were 7.38 seconds and 0.11 seconds, respectively) and by 6 professors in a reverberant room, a semi-reverberant room, and an anechoic room ($T_{30,\ 0.5–2\ kHz}$ were 3.51 seconds, 1.73 seconds, and 0.05 seconds, respectively). Although the differences detected by the authors did not reach significance, they found a tendency for speakers in both groups to increase D_{t_p} with the increase in reverberation.

Several studies have analyzed the relationship between voice acoustical parameters and vocal fatigue. Vocal fatigue can be related to *laryngeal muscle fatigue* and *laryngeal tissue fatigue*. Laryngeal muscle fatigue, which can cause tension in the vocal folds, is caused by depletion or accumulation of biochemical substances in the muscle fibers. Laryngeal tissue fatigue takes place in non-muscular tissue layers (epithelium, superficial, and intermediate layers of the lamina propria) and is caused by changes in molecular structure that result from mechanical loading and unloading.[5] Fundamental frequency and f_0 standard deviation have been found to increase over the course of a work day, as reported by Rantala et al.[6] They analyzed recordings of 33 female teachers during the first and the last lessons on a normal workday. Each lesson had a duration of 35–45 minutes, whereas the workday was 5 hours long. They divided the teachers into two categories: subjects with many voice complaints and subjects with few vocal complaints. The results of the study indicated that some voice features changed during the working day, even if these changes were not monotonic. The most uniform changes were seen in f_0, which increased toward the end of the working day (9.7 Hz, P value < 0.001). The magnitude of the f_0 increase was larger in the subgroup with few vocal complaints (12.8 Hz, P value < 0.001). The f_0 standard deviation showed a similar tendency.

The first aim of the present study was to analyze the effect of the acoustical environment on time dose and f_0, while taking into account the effect of short-term vocal fatigue. Based on the literature results, it has been hypothesized that f_0 means and standard deviations will increase under less reverberant conditions and when the voice becomes fatigued, whereas phonation time will increase under more reverberant conditions.

Based on the same experiment, Bottalico et al[7] reported the effects of room acoustics, voice style (corresponding to normal and raised levels) and short-term vocal fatigue on sound pressure level centered per subject (ΔSPL), and self-reported vocal effort, control, comfort, and clarity. The second aim of the current study was to predict self-reported vocal effort from objective measurements, combining the results of the voice parameters analyzed in this study with the results from Bottalico et al.[7] Based on the standard ISO 9921,[8] vocal effort can be quantified by means of voice SPL. However, it has been hypothesized that other vocal parameters should also be considered to better predict self-reported vocal effort.

EXPERIMENTAL METHOD

The speech of 20 seated talkers was recorded in three different rooms in the presence of artificial babble noise, with and without acrylic glass panels at 0.5 m from the subjects' mouths. More details on and the rationale of the experimental method are given in Bottalico et al.[7] Speech signals were processed to calculate measures of phonation time (D_{t_p}) and f_0.

Subjects, instructions, and equipment

Ethics approval for the experiment was granted by the Michigan State University Human Research Protection Program (IRB 13–1149). Twenty students, comprised of 10 males and 10 females, participated in the experiment. All subjects were aged between 18 and 30 years (mean age 20.8 years), were nonsmoking, and had self-reported normal speech and hearing.

The subjects were instructed to read a text for approximately 30 seconds in duration in the presence of artificial babble noise, with and without acrylic glass panels at 0.5 m from the subjects' mouths. Two different speech styles were used: normal and loud. The instructions given for the styles were as follows: normal: "Speak in your normal voice"; loud: "Imagine you are in a classroom and you want to be heard by all of the children."

The subjects were recorded in three different rooms: an anechoic room, a semi-reverberant room, and a reverberant room. In each room, the subjects were asked to read in four conditions (for a total of 12 tasks): (i) with normal vocal effort and without the presence of the reflective panels; (ii) with loud vocal effort and without the presence of the reflective panels; (iii) with normal vocal effort and in the presence of the reflective panels; and (*iv*) with loud vocal effort and in the presence of the reflective panels. The time separating these tasks was between 15 and 30 seconds. The experimental setup is shown in Figure 1. With the aim of an equal distribution of vocal fatigue (throughout all of) the tasks across subjects and to avoid any other confounding effects of order of administration, the order of administration of the tasks was randomized. With the aim to quantify possible effect of vocal fatigue, the chronological order of tasks administration, which was different for each subject, was considered in the analysis.

Each subject answered several questions after each task. In particular, subjects were asked: "How effortful was it to speak in this condition?" Subjects responded by making a vertical tick on a continuous horizontal line of 100 mm in length (on a visual analogue scale). The score was measured as the distance of the tick from the left end of the line. The extremes of the scale were "not at all" (left) and "extremely" (right).

Speech was recorded using a head-mounted microphone placed 5–7 cm from the mouth (Glottal Enterprises M80, Glottal Enterprises, Syracuse, NY). The microphone was connected to a PC via an external sound board (Scarlett 2i4 Focusrite, High

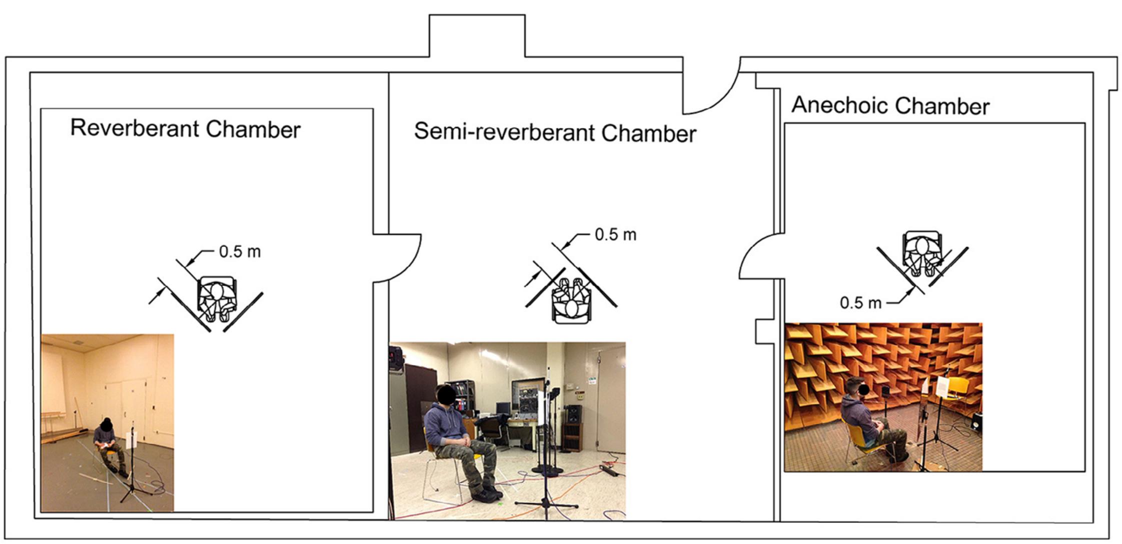

FIGURE 1. Experimental setup during the experiment.