Audio Engineering Society

Convention Paper

Presented at the 131st Convention
2011 October 20–23 New York, NY, USA

Effect of whole-body vibration on speech. Part II: effect on intelligibility.

Durand R. Begault

Human Systems Integration Division (Code TH), NASA Ames Research Center, Moffett Field, CA

Durand.R.Begault@nasa.gov

ABSTRACT

The effect on speech intelligibility was measured for speech where talkers reading Diagnostic Rhyme Test material were exposed to 0.7 g whole body vibration to simulate space vehicle launch. Across all talkers, the effect of vibration was to degrade the percentage of correctly transcribed words from 83% to 74%. The magnitude of the effect of vibration on speech communication varies between individuals, for both talkers and listeners. A “worst case” scenario for intelligibility would be the most “sensitive” listener hearing the most “sensitive” talker; one participant’s intelligibility was reduced by 26% (97% to 71%) for one of the talkers.

1. INTRODUCTION

A set of investigations to characterize the effects of whole-body vibration on speech communications and possible mitigation approaches has been underway at NASA Ames Research Center since 2009, under the colloquial name “Vibrovox.” Part I of this paper published in 2009, titled “Stimuli recording and speech analysis” [1], investigated the effect of 0.5 and 0.7 g whole body vibration on speech production of words, using a Diagnostic Rhyme Test (DRT) word list as a corpus [2, 3]. Six talkers were recorded in that study using a specially designed chair and vibration platform. The effect of the vibration was a very pronounced vocal “shakiness” related to the excitation frequency of 12 Hz, with both amplitude and frequency modulation effects observed in the spectrographic and fundamental frequency (F_0) analyses of the speech. The reader is referred to this companion paper for details of the stimuli used here, and for details on the hardware and methods used for imparting vibration.

In space flight operations, the intelligibility of radio communications between flight deck and ground control is of critical concern, particularly during the launch phase of flight. Requirements for speech intelligibility, including radio communications, are mandated by NASA’s Human-Systems Integration Requirements [4] to provide a level equivalent to a 90% word identification rate. Under conditions of extreme acceleration and vibration during launch scenarios, this goal may not be met for air-ground communications due to adverse impacts to the vocal production system. Effective speech communications are particularly necessary during launch since other means of communication, such as manual operation of switches or keypads, are severely impaired if not impossible.

This report addresses the relative effects of vibration on speech intelligibility for a set of participants listening recordings of an unfamiliar set of talkers, with the talkers exposed to 12 Hz, 0.7 g (zero-peak) vibration. The paradigm is concerned with the communication channel from a speaking crewmember (exposed to vibration) and a listener in a ground control scenario (not exposed to vibration). The talkers in this investigation represent the 'crew member' portion of this communication channel, and the listeners the 'ground control' portion of the channel.

2. STIMULI

Nine participants were presented with previously recorded material from five of the six talkers in [1], recorded under no vibration and 12 Hz, 0.7 g vibration conditions. The vibration in the previous study was effected by a fixed-base vibrating chair platform used in visual display vibration studies [5]. The 192 stimuli words of the Diagnostic Rhyme Test (DRT) described in [2, 3] were read by the participants from four successively placed 10 x 20 inch panels, each containing 48 words printed in 36-point Times Roman. The direction of acceleration was along the *x*-axis, i.e. from the rear to the front of the body.

The participants in this current study were seated in a soundproof audiometric booth, with stimuli provided at approximately 60 dB SPL via circumaural headphones (Sennheiser HD 595). The experimental blocks were continuous, with stimuli randomized between vibrated and non-vibrated speech conditions using custom software.

3. METHODOLOGY

Two separate study protocols were used to measure comparative intelligibility for vibrated and non-vibrated speech.

For the first protocol, participants were given sequential aural presentations of a single word from a 384-word list that was formed from the DRT speech material. No visual reference or training was given for the word list. Participants were asked to transcribe the word heard by typing it into a computer display. Spelling errors or phonetically ambiguous words (e.g., "cheap, cheep") were corrected during post-analysis. A total of 3,456 responses were gathered from the nine participants.

For the second protocol, a self-paced, two-alternative forced choice test was run using the DRT protocol outlined in ANSI S3.2 [4]. Data were gathered for 1920 responses from each participant (17,280 responses total). Participants selected one of two words that were presented visually via a hand-held LED display that also served as the response device.

4. RESULTS

Analysis of the results of the first test protocol (ANOVA for talker and condition independent variables) indicated a significant effect of vibration on speech intelligibility (F(1,8) = 141.3, $p < .001$); see Figure 1. Across all talkers, the effect of vibration was to degrade the percentage of correctly transcribed words from 83% to 74%. The average 9% reduction in intelligibility ranged from 6–14% amongst different listeners. Amongst different talkers, the reduction ranged from 5–13% ; see Table I.

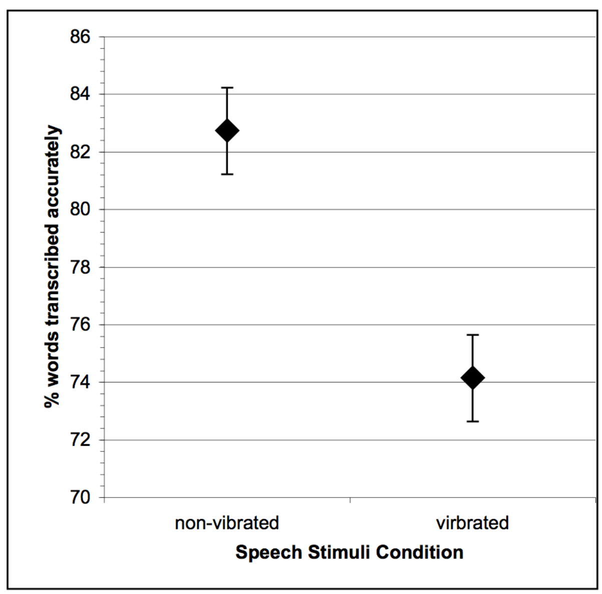

Figure 1. Intelligibility degradation (5 talkers, 9 listeners) for transcription of words for non-vibration and vibration speech stimuli.

Overall, it is possible to conclude that the magnitude of the effect of vibration on speech communication varies between individuals, for both talkers and listeners. A "worst case" scenario for intelligibility would be the most "sensitive" listener hearing the most "sensitive" talker; for example, one participant's intelligibility was reduced by 26% (97% to 71%) for one of the talkers. This type of experiment was the most challenging of the two protocols, since the DRT words are not presented visually to the listener, and the stimuli lacked the cognitive contextual cues that might be used when hearing a sentence containing the word.

t 1 n	17.5
t 1 v	**22.0**
t 2 n	18.3
t 2 v	**30.9**
t 3 n	19.9
t 3 v	**28.6**
t 4 n	13.3
t 4 v	**19.8**
t 5 n	14.4
t 5 v	**27.4**

Table I. Percentage of incorrectly transcribed words across all participants, categorized by talkers t1-t5 under non-vibration (n) and vibration (v) conditions.

A time-accuracy tradeoff analysis indicated the opposite of the usual effect, in that there was a negative correlation between accuracy (correct word transcription) and time to indicate a response. This is partially explained by the fact that many words are in fact longer in duration when spoken under vibration (primarily due to musculoskeletal compensation to produce vowels). It also indicates that faster responses (for most talkers without vibration) were more accurate, while slower responses (characteristic of vibrated talkers) yielded less accurate responses. For the talker most "sensitive" to vibration, responses took 500 ms longer amongst all listeners, with a corresponding 11% decrease in accuracy.

Analysis of the results of the second test protocol (paired t-test) also indicated a significant effect of vibration on speech intelligibility ($t = 17.97$, $p < .001$), reducing the number of correctly identified words from 96% to 94%. The magnitude of the effect of vibration was much lower in this experiment due to the two-alternative forced choice protocol; the participant only had to choose between two words that they could refer to visually. In many cases, the carrier information for allowing discrimination was contained in the initial consonant (e.g., "jaws", "gauze"). In the analysis presented in [1], it was found that the vibration had more of an effect on the production of vowels or final consonants. This experiment therefore represents a less challenging type of intelligibility test, perhaps analogous to when a limited vocabulary is known in advance to the listener.

5. DISCUSSION

Overall, these results indicate that speech intelligibility tests as prescribed in current standards for spaceflight communication systems may be insufficient for potentially realistic vibration scenarios. A significant degradation of intelligibility was observed in the results of two different test protocols, and differing levels of sensitivity were observed for both talkers and listeners. The results are not particularly surprising given prior research [e.g., 6-8], but the typical model for intelligibility standards is usually a talker able to clearly enunciate words. Failing to account for talkers under significant vibration may impact safety, particularly for communication under off-nominal conditions that would require departure from a limited vocabulary.

Intelligibility might be further degraded by the presence of constant 3.-G acceleration in conjunction with the 0.7-g *x*-axis modulation. This level is predicted as a maximum during future spacecraft launch. Informal reports from centrifuge studies indicate that the impact of such a constant force makes speech very difficult to produce; it is expected that further degradation will be observed in speech intelligibility experiments using stimuli from talkers exposed to both vibration and constant g.

For nominal operational conditions involving vibration, it is recommended that speech communication from the flight deck should be minimized or set to conform to a limited vocabulary, and that methods of processing the communication to enhance intelligibility should be evaluated. Under off-nominal situations where launch vehicle speech communications might be necessary, e.g., to describe a specific problem using an unlimited vocabulary, the impact of these results may require solutions to "correct" the audio signal. One approach currently under development by the author involves post-processing of the audio signal into "corrected speech" to minimize both the frequency and amplitude modulation distortion that results from vibration. Verification of this approach would involve conducting the same studies described here but using the "corrected speech", to determine if intelligibility can be restored to the baseline "no vibration" condition.

6. ACKNOWLEDGEMENTS

This work was supported by NASA's Space Human Factors Engineering Project (SHFE) Directed Research Project Plan (DRPP) on Information Presentation. Many thanks to Mark R. Anderson, Bernard D. Adelstein, the research staff of the Advanced Controls and Displays Group at NASA Ames Research Center, and all of the experimental participants.

7. REFERENCES

[1] Begault, D. R. "Effect of Whole-Body Vibration on Speech. Part I: Stimuli Recording and Speech Analysis" *Audio Eng Soc 127th Convention, NY* Paper 7820: October 2009.

[2] ANSI S3.2-1989 (rev. 1999). Method for measuring the intelligibility of speech over communication systems. American National Standards Institute (1999).

[3] Voiers, W. D. "Evaluating Processed Speech Using the Diagnostic Rhyme Test". *Speech Technology* (Jan/Feb 1983) 30-39 (1983).

[4] National Aeronautics and Space Administration. *Constellation Program Human-Systems Integration Requirements*. Document CxP 70024 (2006).

[5] Adelstein, B. D.; Beutter, B. R.; Kaiser, M. K.; McCann, R. S.; Stone, L. S.; Anderson, M. R.; Renema, F.; and Paloski W. H. Influence of Combiner Whole-Body Vibration Plus G-Loading on Visual Performance. NASA_TM-2009-215386 (2009).

[6] Griffin, M. J. *Handbook of Human Vibration*. London: Elsevier Ltd. (1990).

[7] Nixon, C. W. "Influence of selected vibrations on speech I. Range of 10 cps to 50 cps." *The Journal of Auditory Research* 2: 247-266 (1962).

[8] Nixon, C. W. and Sommer, H. C. Influence of selected vibrations on speech III. Range of 6 cps to 20 cps for semi-supine talkers. *Aerospace Medicine* 34:1012-1017 (1963).