COMPARISONS OF TASK EFFICIENCY IN FACE-TO-FACE, DOLBY® VOICE™ TECHNOLOGY, AND TRADITIONAL CONFERENCE SYSTEM COMMUNICATIONS

Introduction

Dolby® Voice™, the next generation of audio conferencing offering high-quality audio with spatial rendering of the participants, was launched in 2013. It is evident that systems like Dolby Voice with the Dolby Conference Phone reduce the gap between face-to-face conversations and online meetings, and it is important to standardize appropriate subjective test methods for evaluating just how effective this approach can be. A number of studies have already demonstrated the limitations of existing standardized subjective test methods.[1]

Effective performance of communication systems relies on a number of factors, such as double talk performance (full duplex), loudness balance among participants, talker separation, and so on. In order to evaluate these factors in a conversational test, multiparty tasks that stress the systems under test are desirable. Therefore, highly interactive tasks such as the one described in ITU-T Rec. P.805, Appendix VIII, are preferred.[2, 3] These tests simulate real-world use cases in which remote workers are engaged in collaborative discussions.

The Experiment

The experiment involves the participation of four or more subjects. Grouped into two teams, with two persons per team, the participants complete tasks concurrently and on the same communication link. The task consists of one member of each team, the "reader," communicating a sequence of letters or words to the partner(s) or the "responder(s)." The objective is to communicate the highest number of sequences within the given time. Each team evolves at its own pace, so at the end of task, each team will likely have a different number of sequence iterations.

During a task, the designated readers pronounce a sequence of six letters or words. Their partners enter the sequence they hear via a graphical interface. The task for a given team continues with a new sequence if the submitted response is correct. If the response is incorrect, the responder is visually alerted and invited to submit a corrected sequence.

The test results in an objective performance metric, which is the number of letter and word sequences correctly relayed for each team. In addition, participants responded to a P.805-style survey addressing quality and effort questions at the end of each session.

Systems under Test

In this experiment, three different communication systems were used:

- **Face-to-face:** Face-to-face provides a reference condition. All participants are seated around a conference table, and participants on the same team face each other.
- **Dolby Voice:**A low-bitrate wideband conferencing system with full-duplex capabilities and spatial rendering. Participants communicate using a laptop computer and business-grade stereo headset as remote participants, and using the Dolby Conference Phone while in the conference room.
- **Traditional conference system:** A mono, standards-based conference bridge. Participants communicate using a laptop computer and business-grade stereo headset as remote participants and using a Polycom IP7000 while in the conference room.

All of the systems were noise-free links and provide acceptable end-to-end delay.

Test Setup

The subjective test was conducted in a professional-grade subjective test facility, and the environmental noise was equivalent to that found in a quiet office.

The face-to-face tasks were conducted in a single booth the size of a conference room. To test the two conferencing systems, one participant from each team was seated in an individual booth (remote participants) and one participant from each team was seated in the conference room (conference room participants). The remote participants connected to the conferencing systems with a laptop computer and the same set of business-grade stereo headsets. The participants were not told which systems they were connected to, and the transitions from one conferencing system to another were "transparent"; that is, the line was muted until the audio conference was set up.

Each participant received instructions and entered responses via a simple graphical interface on a tablet that each person carried between the conference room booth and the individual booths. The tablets were remotely monitored for response validation and time control.

Subjects

The test was performed with seven groups of four participants who were not audio or telecommunications professionals and who did not know each other.

All participants were native English speakers. Genders and age groups were well represented among the groups, and the teams remained the same throughout the entire test.

The test administrator provided instructions to the participants and conducted a training session face-to-face with the participants to familiarize them with the tasks and interface.

Furthermore, the test administrator led a brief and informal discussion between participants after each connection to a different system, which allowed each team's reader and responder to get used to each other's voice before the start of the tasks.

Experimental Results

Each test consisted of two to four rounds of tasks on all of the communication systems and took three hours to complete. The systems were presented in different order for each of the rounds; face-to-face was always presented first, and the two audio conferencing systems were presented in random order.

Measure of Correct Responses

Because the tasks are time limited, the number of correct answers provides a measure of the effectiveness of the system; the more correct responses are received, the more effective is the communication over the system.

Participants achieved the highest number of correct responses in the face-to-face configuration. The average number of correct answers for each system, along with 95 percent confidence intervals, is given in Figure 1. These results show a statistically significant difference between the three communication systems, with the ranking by correct responses as follows: (1) face-to-face, (2) Dolby Voice conferencing, and (3) the traditional conferencing system.

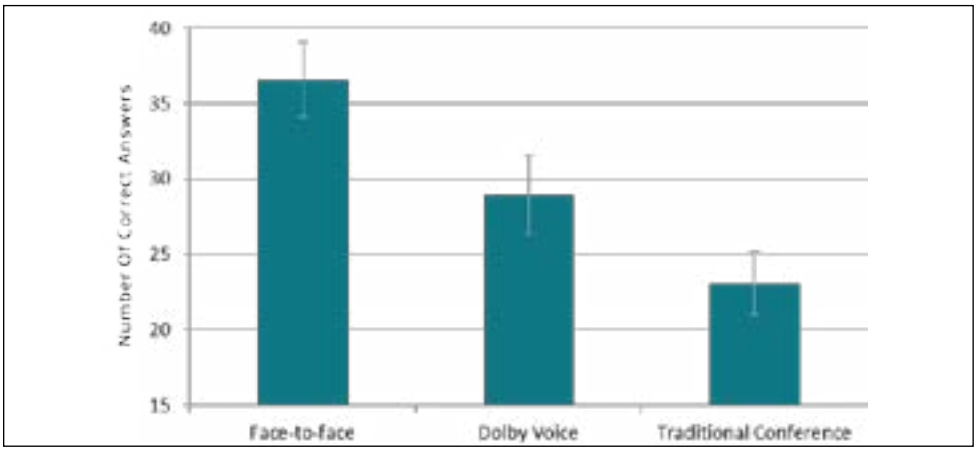

Figure 1. Average number of correct answers per system. Each bar shows a confidence interval that is bounded by a gray vertical; bounds are unique to each test.

Effectiveness Relative to Face-to-Face Meeting

The face-to-face configuration is, as one might expect, the best scenario for completing the tasks and forms a good reference point to compare conferencing systems. The system effectiveness relative to face-to-face communication can be calculated by normalizing the average number of correct answers within a group to the highest score achieved by this group on any of the systems. The average system effectiveness is provided in Figure 2 and shows a clear separation between the systems under test. The variation in performance of each of the systems is statistically significant.

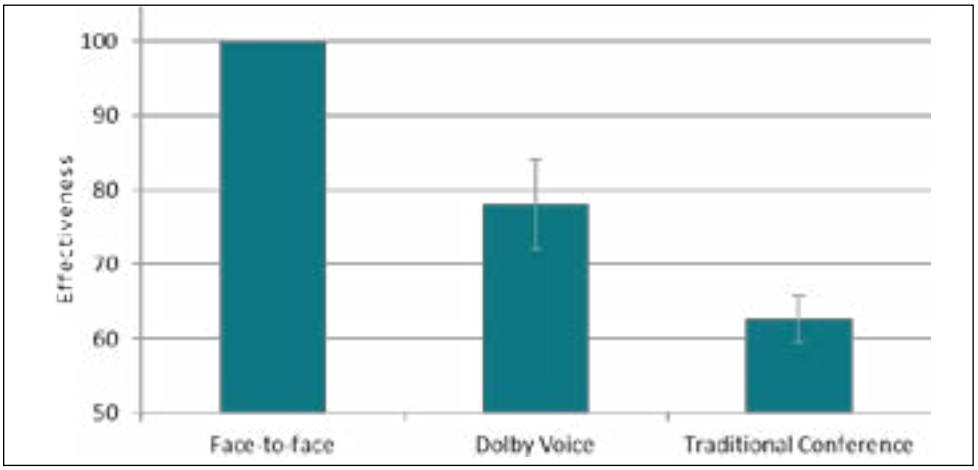

Figure 2. Average system effectiveness. Each bar shows a confidence interval that is bounded by a gray vertical; bounds are unique to each test.

The system effectiveness could be viewed from a different angle. Let's assume that a task in this test requires a number of sequence iterations to be completed instead of being time limited. We can infer from the system-effectiveness results that if a task takes 60 seconds to complete face-to-face, 80 seconds would be needed to complete it on the Dolby Voice system, compared to 134 seconds on a traditional conference system.

P.805-Style Subjective Assessment

At the end of each set of tasks on each of the three systems under test, participants were requested to provide their assessment of the system on the following dimensions:

- Opinion of the connection, rated on a five-point quality scale
- Difficulty in talking or hearing, rated yes or no
- Sound quality of the partner's voice, rated on a five-point degradation scale
- Ease of understanding what the partner said, rated on a five-point effort scale
- Ease of recognizing the partner's voice, rated on a five-point effort scale
- Effort needed to converse back and forth, rated on a five-point effort scale

The percentage of responses reporting difficulties in hearing or talking to the partners are provided in Figure 3, and the rank order of the systems once again is in agreement with the rest of this study.

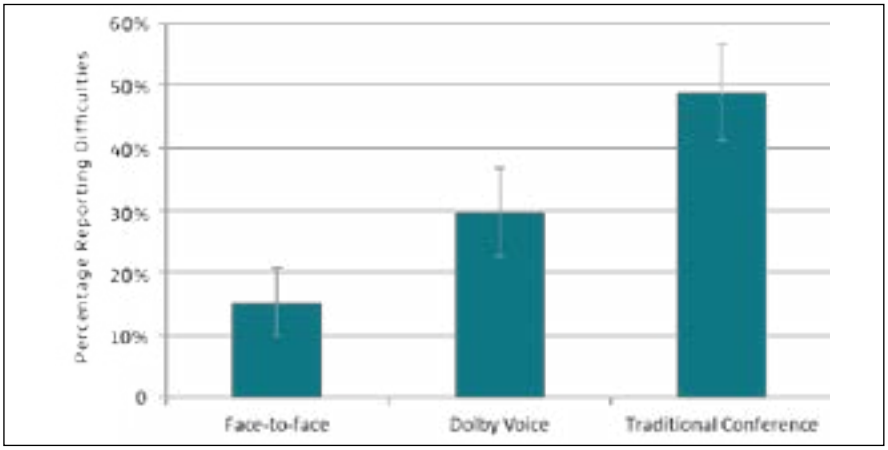

Figure 3. Percentage of participants reporting hearing or talking difficulties. Each bar shows a confidence interval that is bounded by a gray vertical; bounds are unique to each test.

Conclusions

This white paper presents the results of a task efficiency test procedure that Dolby Laboratories is investigating. The method provides quantitative metrics to assess the performance of conferencing systems with respect to task efficiency in multiparty online meetings. The results show that the measurements and P.805-style subjective assessments are highly correlated, which indicates that the proposed tasks are appropriate for this experiment.

The methodology also presents the advantage of comparing the performance of audio conferencing systems with face-to-face meetings, an idealized reference system that is easily replicable across different test laboratories.

Three different meeting configurations were considered in this study, and the results show significant differences in performance. It is clear that Dolby Voice as an end-to-end solution including the Dolby Conference Phone significantly reduces the gap between traditional audio conferencing and face-to-face meetings

References

[1] Rapporteurs for Question 10/12, "Status Report of Question 10/12," *TD* 86rev2, March 2013.

[2] ITU-T Rec. P.805, "Subjective Evaluation of Conversational Quality," April 2007.

[3] Kitawaki, N., and Itoh, K., "Pure Delay Effects on Speech Quality in Telecommunications," *IEEE Journal on Selected Areas in Communications*, vol. 9, no. 4 (1991): 586-593. (Issues are available online to IEEE members.).

DOLBY **Dolby Laboratories, Inc.** | 100 Potrero Avenue, San Francisco, CA 94103-4813 USA | T 415-558-0200 F 415-645-4000 | **dolby.com**