DO BYSTANDERS AND DIALOG PARTICIPANTS DIFFER IN PREFERENCES FOR TELECOMMUNICATIONS CHANNELS?

NIGEL G. WARD, ANAIS G. RIVERA, ALEJANDRO VEGA
University of Texas at El Paso
500 West University Avenue
El Paso, Texas 79968, USA
{nigelward@acm.org, agrivera@gmail.com, avega5@miners.utep.edu}

Abstract. Mobile telephone conversations in public places are often annoying to bystanders. Previous work has focused on the psychological and social causes for this (Monk 2004b, Ling 2004), but has not examined the possible role of properties of the communication channel. We hypothesized that some of the annoyance could be explained if bystander preferences differ from talker preferences. If this is this true, it will be possible to develop telephony infrastructure which enables users to have mobile conversations in public places without annoying others. This paper report a series of preliminary studies, done to find an experimental method that would enable the demonstration of the existence of divergent preferences. The strategy was to have both talkers and bystanders judge conversations across a moderate-noise low-delay line and across a low-noise moderate-delay line. No clear tendency was found; we were unable to conclude that a preference difference exists.

1. Introduction: Why are Mobile Telephone Conversations Annoying?

It is often remarked that mobile phone conversations are more annoying, to those who overhear them, than normal conversations. Many possible reasons for this exist (Ward et. al., 2007), and two have been experimentally studied: loudness and the lack of a audible interlocutor. Monk et. al. (2004a) have shown that, although mobile conversations may be louder than face-to-face conversations, and although loudness does have an effect on annoyance, mobile phone conversations are more annoying even when no louder than face-to-face conversations. Monk et. al. (2004b) have further shown that hearing only one side of a conversation can induce a psychological ``need to listen" that can make it seem more intrusive; however it is not known how much of the perceived annoyance can be explained by this effect. Several other likely contributing factors have been identified (Ling 2004); however these have not been subjected to experimental evaluation. In this report we describe a series of experiments exploring the possibility of reasons of a different type.

We hypothesize that (1) for telecommunication channels bystander preferences differ from user preferences. Today's telecommunications infrastructure is carefully engineered to optimize the benefit to the users. In particular the voice coding methods and transmission methods are chosen to maximize the user's perception of channel quality. The results of many human factors studies have culminated in a quantitative model that accurately predicts how perceived channel quality depends on noise, delay, and echos (ITU 2005). In mobile telephony, however, conversations commonly have participants other than the users, namely bystanders in the vicinity. These participants also have perceptions of benefit and cost which cannot be ignored.

It seems likely that bystanders are indirectly sensitive to and affected by the channel properties in mobile telephony, and that their preferences may differ from those of talkers. The differences could arise in several ways. Line delay is known to upset conversation dynamics (Krauss and Bricker, 1967; Karis, 1991) and lead to more talking-over the other person, which could cause the talkers to compensate by using more explicit turn-taking prosodic signals. This may tend to attract the attention of bystanders. Thus we originally hypothesized specifically that (2) bystanders dislike channel delay more than do talkers. Regarding line delay, there is the additional possibility that the resulting disruption of the familiar on-off patterns of dialog may make the overheard voices harder to ignore. On the other hand, there are also reasons to suspect that bystanders might dislike channel noise more than do talkers. For example, noise on the line may cause the talkers to speak louder or more stridently (the Lombard effect), and these voice properties may be more annoying to bystanders. Noise on the line might also cause talkers to struggle to hear the other party, increasing their cognitive load and decreasing their ability to modulate their voice to make it inoffensive to bystanders.

In either case, if differences do exist, then current models are inadequate as guides for the design of systems which may be used in the presence of third parties. This may be true not just for the design of mobile telephony systems, but also for systems which enable users to have spoken language interaction with computers, which is an important and growing way to access information while mobile. If such differences exist, then mitigating the annoyance becomes a technical challenge rather than a social one, and improvements in telephony infrastructure could alleviate the need for crude strategies such as banning all mobile phone conversations in trains, libraries, etc. for the sake of bystanders. Thus it is important to determine whether and how the preferences of bystanders differ from those of users.

In a preliminary study in our laboratory, David Ponevac (2006) explored perceptions of channel delay. The most relevant experiment involved 9 dialog pairs, presented to a grand total of 23 bystanders. Each dialog pair was between the two people, one co-located with the bystanders in the classroom and the other in a different room. The only difference was in the communications channel: the first of each dialog pair involved 150ms delay and the second 250ms. Ponevac found that 14 of the 23 bystanders (61%) considered the channel with more delay to be more annoying, but only 8 of the 18 talkers (44%) did. This is in line with both hypotheses 1 and 2. However the lack of balance in the order of presentation and lack of statistical significance led us to investigate further.

2. Research Strategy

We accordingly set out to find whether bystanders' preferences do differ from those of users. As not all mobile conversations are annoying, our first task was, necessarily, to find a situation in which we could demonstrate that a difference exists. We also needed to find an appropriate experimental method. Our basic strategy was to look for a pair of telecommunications channels where talkers would reliably prefer one but bystanders the other. At the same time, we were looking for a suitable experimental method, trying various conversation scenarios, various ways of presenting the dialogs to the bystanders, and various ways of eliciting preferences. Since echo is probably not a major problem in most mobile dialogs, we focused on the other two main determinants of user impressions, delay and noise. The experiments all involved two channels, one with moderate noise and low delay, Cn, and one with low noise and moderate delay, Cd, as illustrated in Figure 1. Note that the bystander hears direct or directly recorded audio, not audio that has been transmitted.

The low noise channels generally used the G.711 codec, which is familiar from landline telephony, although in one experiment we used CD-quality audio. The high-noise channels used the GSM-FR codec, and were built on top of the "toast" program (Deneger, 1992). Such small variations in noise are not always identifiable by the untrained ear, especially if the users' attention is not drawn to them, however they do tend to affect users' judgments of line quality.

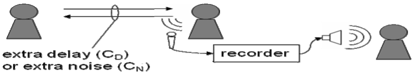

Figure 1: Experiment Concept.

The low-delay software introduced a delay of about 150 milliseconds. The high-delay versions used double-buffering to introduce various amounts of delay, most commonly 350 milliseconds. Both are within the realm of common experience; measurements of line delay in calls to and among mobile phones in the El Paso area ranged from 64 ms to 386 ms, depending on the mobile phone providers. Delays were measured by measuring the interval between the time a source click was produced beside the microphone and the time its delayed version was produced by the headphone.

Manipulating the channel properties of actual mobile telephones being infeasible, we simulated the channels in software. Speech from one talker was captured from the microphone, processed, and sent to the headset for the other talker. A second computer similarly processed speech in the other direction. The programs, one for each channel, ran on otherwise idle Linux boxes. At each time the channels were the same, thus each speaker heard the other's voice in the same way.

Each speaker wore a headset consisting of a microphone and clamshell headphones. Each speaker also had a desk microphone in front of him. The audio from the desk microphone, modified, was sent to the headphone of the person in the other room. The audio from the headset microphone was fed into a digital recorder for later use. Clamshell headphones were needed so that sound did not leak back into the microphone and cause an echo; it was still necessary to reduce the gain to prevent this. The speakers were seated in medium-sized basement offices, each carpeted and containing computers, desks, chairs, and tables. The offices were separated by a foyer and two closed doors to prevent direct sound propagation.

The subjects were recruited in various ways. Most were students from the Computer Science subject pool, and were compensated only by class credit. However some were recruited by fliers or word of mouth; these were generally compensated, either with pizza and drinks or with $10. The pool included many bilingual speakers, mostly Spanish-English bilinguals, however most of the time we asked them to converse in English. The talkers were generally recruited in pairs, and so they were mostly friends or acquaintances. The order of channels was rotated. Each pair of dialogs was between the same two people, in the same situation, with only the channel changed.

These dialogs were then used as stimuli for listeners, also recruited from the subject pool. The instructions informed them of our interest in determining how distracting various kinds of conversations are. We generally tried to give them a distracter task or asked them not to pay particular attention to the stimuli, since our interest was in the perceptions of people subjected to conversations while they're doing other things. This series of experiments was accepted by the UTEP Institutional Review Board, and granted exemption #2513.

3. Experiments and Results

We tried a succession of experimental methods (Ward et. al., 2007). This section summarizes.

3.1 FIRST APPROACH

In this attempt we gave each talker four cards, each containing seven numbers which had between one and seven digits. Each "dialog" consisted of the subjects reading numbers to each other, switching off at each number. This was a variant of Kitawsaki and Itoh's (1991) Task 1, with the length of the numbers being varied because we thought this would make it more necessary for the speakers to use turn-taking cues, specifically to signal prosodically when each number was complete.

Subjects used both Cn (GSM with 150ms delay) and Cd (no GSM and 300ms delay). We gave them a survey that asked them to compare them on dimensions such as channel quality, the attentiveness of the dialog partner, and the likelihood of being annoying to bystanders, with a forced choice between the first dialog or the second one. Then we had them listen to a pair of recordings of exchanges in both conditions and take another survey to rate which was more annoying, intrusive, noticeable, etc (Monk, 2004a). The recordings we asked them to judge were from previous sessions. There seemed to be some tendency for the talkers to find Cd worse than Cn, whereas the bystanders had no clear preference. However the tendency was very weak.

3.2 SECOND APPROACH

As a more realistic situation we had two participants have a normal unconstrained conversation while several bystanders in one of the rooms ate pizza. After the conversations we asked the participants, both the talkers and bystanders, to rate which conversation was more annoying etc. They also listened to recordings from a previous session and rated those. Again the differences between speaker and bystander preferences were weak at best. Varying aspects of the protocol, such as the distractor tasks and the proximity between the bystanders and the talker, again did not yield clear differences.

3.3 THIRD APPROACH

At this point we began to wonder whether our original hypothesis (2) was correct, as much of the evidence seemed to be in the other direction, suggesting that Cn was relatively more annoying to bystanders. We even began to question hypothesis (1), that is, we began to wonder whether there was an effect to find, and if there was one, whether our methods were sufficient to reveal it. We decided to try a more extreme manipulation. Recalling the observation that the effects of line delay are less noticeable in dialog with long turns (Kitawaki and Itoh, 1991), we decided to accentuate the effects of delay by adopting their Task 1, in which the subjects read from two lists of single digit numbers, switching off at each digit, attempting to do so as fast as possible. We conducted two experiments of this type; one with the usual 350 ms delay and one with 500ms delay. Even though we had attempted to accentuate the effects of delay, overall both talkers and bystanders still seemed to dislike Cn more.

3.4 FOURTH APPROACH

In the previous experiments we had sought to measure overall impressions of substantial chunks of dialog. In this experiment we wanted to see if focused listening to specific short number sequences across the two conditions would reveal a clear preference. Using the recordings from the previous attempt we played matching sequences of numbers excised from both conditions to three subjects and asked them to rate their preference after each pair.

Here there was a tendency for the subjects, here all acting as bystanders, to dislike Cn more, with the tendency being pronounced for Subject 1. The experimenters also listened to the samples, to see whether they could perceive any consistent difference in annoyingness or other qualities between the two channels; here there was poor agreement and no overall tendency. We also did one run in which the characteristics of the Cd channel were exaggerated: the delay was increased to 800ms and the audio quality to CD-quality: 44.1kHz, 16 bit. Here we expected to see a truly obvious effect, but, as usual, the judgments were not extreme.

3.5 FIFTH APPROACH

In a last attempt, the protocol was kept the same but the surveys were redesigned. Instead of a forced choice the survey used scales labeled from 0-10 (Möller 2000), on which subjects placed a mark at the place corresponding to their rating of the channel they were using or overhearing.

The averages suggest that talkers preferred Cd but bystanders preferred Cn, as seen in Figure 2. However the variation was clearly large, and a matched-pair two-tailed t-test over the 16 subjects confirms that this tendency was weak ($P \sim 0.4$).

We also noted that subjects were fairly enthusiastic about diagnosing the properties of the two channels, which most were able to do correctly. This raises the possibility that their judgments were affected by this, for example, they may have noticed the delay, believed that delay was annoying, and rated Cd poorly for that reason, rather than using their direct impressions.

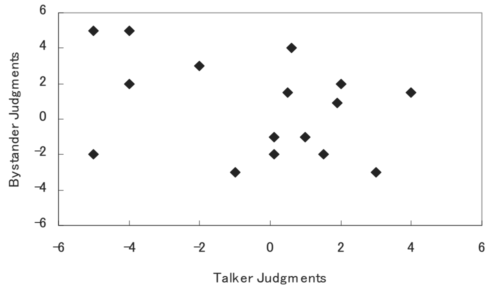

Figure 2. Subject's preferences for Cn over Cd. Each dot represents one subject's judgments, with the difference between their judgments of Cn and Cd on the x-axis when perceived as talkers and on the y-axis when perceived as bystanders.

4. Summary

Methodologically, we learned a few things. Even extreme manipulations, such as boosting the delay, using Kitawaki's Task 1, and having judges devote full attention to listening, did not have the expected clear effects. Subjects are clearly affected by the content of the conversation, but they may be able to factor this out if requested. Subjects who are distant from the conversant appear to be less affected. Subjects were willing to rate the annoyingness even of prerecorded voices, although we do not know how well this reflects real-world annoyance.

The results clearly did not fall into any consistent pattern. We did a meta-analysis across various of these experiments, using the question-pair which seemed most likely to reveal a difference in preference, namely "Which line seemed to be worse quality?" for the talkers and "Which dialog was more annoying?" for the bystanders. For each of the 59 dialog stimuli, we compared the talker's rating and the average rating of all the bystanders. The bystanders average was 1.42 whereas the talkers average was 1.47, suggesting that bystanders disliked Cn more than did the talkers, contrary to hypothesis 2. However this effect was weak at best, and a matched-pairs two-tailed t-test confirmed that this was far from significant. Thus our tentative answer to the motivating question, "do bystanders and dialog participants differ in preferences for telecommunications channels?" is, contrary to hypothesis 1, “probably not significantly”. The main causes of the perceived annoyance probably lie elsewhere.

Acknowledgments

We thank David Novick and Sebastian Möller for insight and discussion, Udit Sajjanhar for the initial implementation, David Ponevac for the first pilot studies, and Ana I. Schwartz for comments on the experiment design. This work was supported in part by the National Science Foundation under awards EIA-0080940 and IIS-0415150 and by DARPA.

References

Deneger: 1992, GSM 06.10 *lossy speech compression library*, Jutta Degener, [online] http://kbs.cs.tu-berlin.de/~jutta/toast.html (accessed October 2006).

ITU: 2005, The E-model, a computational model for use in transmission planning. ITU-T Recommendation G.107, *International Telecommunication Union,* 1998, revised 2005.

Karis: 1991, Evaluating Transmission Quality in Mobile Telecommunications Systems using Conversation Tests. Demetrios Karis, *in Proceedings of the Human Factors Society 35th Annual Meeting*, pp 217-221.

Kitawaki and Itoh: 1991, Delay Effects on Speech Quality in Telecommunications, Nobuhiko Kitawaki and Kenzo Itoh, *IEEE Journal on Selected Areas in Communications*, 9,586-593.

Krauss & Bricker: 1967, Effects of Transmission Delay and Access Delay on the Efficiency of Verbal Communication, Robert M. Krauss and Peter D. Bricker, *Journal of the Acoustical Society of America*, **41** (2), 286-292.

Ling, R.: 2004, *The Mobile Connection: The Cell Phone's Impact on Society*, Morgan Kaufmann.

Möller: 2000, *Assessment and Prediction of Speech Quality in Telecommunications*, Sebastian Möller, Springer, 2000.

Monk, A. et. al.: 2004a, Why are Mobile Phones Annoying? Andrew Monk, Jenni Carroll, Sarah Parker and Mark Blythe, Behavior and Information Technology, 23, pp 33-41.

Monk, A. et. al.: 2004b, Hearing only one side of normal and mobile phone conversations. Andrew Monk, Evi Fellas and Eleanor Ley. *Behavior and Information Technology*, 23, 301-305.

Ponevac, D.: 2006, *Effects of Line Delay and Novice on Bystanders' Perceptions of Mobile Phone Conversations,* University of Texas at El Paso, Department of Computer Science, Masters Thesis.

Ward, N. et. al. 2007, *Reasons why Mobile Telephone Conversations may be Annoying: Considerations and Pilot Studies*, Nigel G. Ward, Anais G. Rivera, Alejandro Vega. Technical Report UTEP-CS-07-60, University of Texas at El Paso, Department of Computer Science.