

Speech-Based Text Correction Patterns in Noisy Environment

Ladislav Kunc, Tomáš Macek, Martin Labský, and Jan Kleindienst

IBM Prague R&D Lab
V Parku 2294/4, 148 00 Prague 4
{ladislav_kunc1,tomas_macek,martin.labsky,
jan_kleindienst}@cz.ibm.com

Abstract. We present a study focused on observation of methods of dictation and error correction between humans in a noisy environment. The purpose of this study is to gain insight to natural communication patterns which can then be applied to human – machine interaction. We asked 10 subjects to conduct the standard Lane Change Test (LCT) while dictating messages to a human counterpart who had to note down the message texts. Both parties were located in separate rooms and communicated over Skype. Both were exposed to varying types and levels of noise, which made their communication difficult and forced the subjects to deal with misunderstandings. Dictation of both short and longer messages was tested. We observed how the subjects behaved and we analyzed their communication patterns. We identified and described more than 20 elementary observations related to communication techniques such as synchronization and grounding of parties, error checking and error correction. We also report frequencies of use for each communication pattern and provide basic characteristics of driving distraction during the test.

1 Introduction

Various communication, navigation and entertainment systems are becoming a part of our everyday in-car experience. Managing them without compromising safety becomes an issue. Appropriate choice of user interface techniques tailored for in-car use can be a critical factor in minimizing driving distraction.

Although many in-car devices nowadays primarily use various forms of not yet standardized controls (rotary knobs, joysticks or touch screens), speech UIs seem to be very promising. Correct design of speech interfaces for cars is however a relatively complex task. It requires proper judgment of various aspects including limited short term memory of the user under cognitive load, finding the right way how to benefit from other modalities, allowing fast recovery after UI interaction had to be interrupted by handling road situation and many other aspects. As part of designing a dictation and error correction UI for a car (see e.g. ECOR [2]), one thing which comes to mind is whether we could benefit from systematically observing the ways how humans do the same task themselves. We summarize the knowledge gained during this study in this paper. Although the study does not bring any quite unexpected observations we

found collected experience very beneficial to our main, man-machine interface design task. It helped us gain a good insight into how human dialog is organized and what are the basic tasks and methods used.

Obviously, we do not imply here that communication patterns between the two humans are the same as in human to machine conversation. If the user is aware that the machine is “on the other end” s/he adapts the conversation to the limits of the machine. The purpose of this study is to learn natural communication patterns of the humans for the task of dictation in a car. Deeper understanding of it should help us in future UI design.

2 Related Work

Significant attention was devoted in the past to judging the impact of in-car activities [1]. The Lane change Test (LCT) [3] and subjective tests using questionnaires such as NASA TLX and DALI [6] are examples of popular methods used to assess the impact of various in-car tasks that are secondary to the primary task of driving.

Although electronics systems are more and more abundant in cars, which rightfully causes worries about their impact on driving, communication between the driver and passengers is frequent and hardly can be regulated [5].

Several approaches to designing speech-based UIs for in-car usage were described including menu-based and search-based UIs [8].

Various patterns of text entry correction were elaborated in desktop speech-based solutions [4]. This paper discusses similar strategies for the car environment.

3 Experimental Setup

We decided to collect communication patterns for a constrained scenario where the driver dictates pieces of text over a hands-free phone. The driver’s counterpart, the note-taker, simulates a person sitting on the passenger’s seat who communicates with the driver. Non-verbal communication was not considered and is out of the scope of this paper.

Tests were conducted in laboratory environment. The drivers were using a low-fidelity driving simulator to mimic driving on a highway. The primary task performed was the standard LCT. Driving statistics were recorded and they are reported in Section 6.1. Each driver communicated over Skype with a note-taker located in a different room. The participants did not have any visual contact. The note-taker typed the received text to a computer text editor.

3.1 Distraction by Noise

To introduce communication errors, we exposed both the driver and the note-taker to a pre-recorded noise audio which was mixed with Skype audio in real-time and the mixed result was playing in headphones for both parties.

One can expect (confirmed by tests) that behavior of the user depends on noise level and also on its character. During pilot experiments we used noise of a starting car followed by more quiet periods of running engine. This recording was played in a loop. We observed fast adaptation of the subjects to the character of the noise as they quickly learned to wait for the quiet periods. Although this is a valid usage pattern, real-life types of noise are frequently not of this kind and cannot be always predicted. Therefore the noise used in the experiments presented here was modified to contain a high variety of sounds. A fragment of the utilized noise track is depicted by Fig. 1.

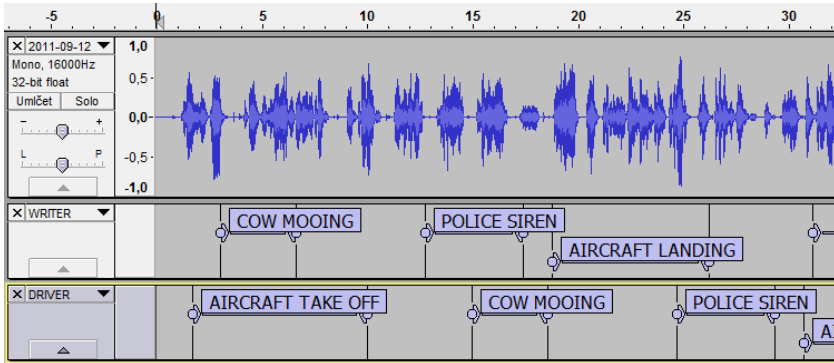


Fig. 1. Noise pattern and alignment of noise tracks for both note-taker and driver

Another lesson learned from the pilot study was about the way how to introduce the noise. Originally we only introduced noise only at the note-taker's side but this did not represent sufficiently the whole variety of natural scenarios. The final tests were conducted with noise added on both sides. Although the same audio was played both to the driver and note-taker, the audio streams were not mutually synchronized so one did not know the noise condition of the other.

Skype communication was also recorded without noise to simplify transcription work needed to analyze the data.

3.2 Dictated Messages

The other important aspect is the correct selection of messages to be dictated. We wanted to have both short and long messages covered in the portfolio of the dictated text due to a suspicion that behavior of the users could differ depending on the message length. One problem was to ensure natural behavior as in normal conditions when the driver dictates messages to a passenger. The method we used was the following. The users were advised to dictate short messages about a selected topic. However the exact formulation was left up to the driver. Regarding the long messages, the drivers were asked to select a topic from their past experience (for example some weekend story) and write the text down prior to the test. The writing exercise ensured that the dictating person does not change the story during the dictation due to some communication problems.

4 Testing Plan

Testing procedure consisted of the following steps:

- Explanation of the test procedure.
- Pre-test questionnaire.
- Learning to use the driving simulator (not recorded).
- First undistracted ride: 3km of LCT driving with no secondary task.
- Dictation of short messages. 2 LCT tests, 3km driving, secondary task is the dictation of semantically prescribed short messages. This part is repeated twice with a different set of messages.
- Dictation of a long message. LCT test, 3km driving, secondary task is the dictation of a previously written long text.
- Second undistracted ride. LCT test, 3km driving, no secondary task.

4.1 The Roles

Each test was conducted under supervision but with minimal intervention of the supervisor with the two participants.

Participants had previous driving experience. To make the most of participants we swapped them so each acted both as the driver and note-taker.

- Driver (D) – drives while performing LCT and dictates message to the note-taker.
- Note-taker (N) – receives the dictated text and records it by typing into a text editor.

5 Evaluation of Results

The test was conducted on 10 individuals. The participants were males – mean age 25.1 (SD 3.8), and they were university students of computer science. The study was conducted in Czech. All participants were native Czech and had Czech cultural background. Before the experiment, each of the participants was instructed verbally on how to proceed with the experiment. The experiment results were anonymized.

By careful analyses of the recorded activities we have identified the following dictation patterns, error correction methods and communication strategies. We state them grouped based on their character. The CHARACTER.NUMBER codes (e.g. C.1, S.1) presented in the next section are references to Fig. 2 (frequency of strategies) and the numbers in parentheses are actual frequencies of each strategy.

5.1 Confirmation Strategies

C.1 Simple acknowledgement (258). When the note-taker is confident, s/he confirms usually only by short confirmation words or sounds (example: “OK”, “Yes”, “Got it”, “Hmm”).

C.2 Acknowledgement by repeating (123). Note-taker confirms what has been understood by repeating the utterance which both confirms that message has been received and provides means to verify the content.

C.3 Request for confirmation (32). A strategy used both by driver and note-taker. The speaker checks if the message has been received (example: “Did you hear me?”).

C.4 Request for confirmation without expected reply (1). It happens that the dictating person asks without real waiting for confirmation (example: “OK?” ...<following dictated text>).

5.2 Synchronization and Communication Channel Handling

S.1 Speeding up or slowing down (30). If the note-taker is not fast enough or dictation is too slow s/he requests slowing down or speeding up by a command (example: “slowly, please!”, or using a confirmation (“I have only got: Buy oranges...”).

S.2 Refinement of the question when no response received (2). If the note-taker does not answer promptly, the driver tends to refine their utterances (example: D: “Barbara” N: long pause D: “Barbara, it’s the English name”).

5.3 Error Prevention

E.1 Ambiguous confirmation (1). Some of the confirmations can be misunderstood and therefore the dictating person adds extra explanation or spelling (example: “Skoda – I mean the car”).

E.2 Active search for quiet intervals (9). The note-taker indicates, on his initiative, the presence of quiet intervals to the driver by short phrases (example: “OK”, “go on”, “now”) which helps to pass the message during favorable noise conditions.

5.4 Error Correction Strategies

R.1 Correction specified by interrogative pronoun (29). The note-taker asks for a specific part which was not understood using an interrogative pronoun question (example: D: “Then you take the mashed potatoes” N: “Take what?”).

R.2 Explanation notes (3). If the note-taker does not understand repeatedly, the driver starts to describe in other words or using explanations (example: “Villon, the French poet”).

Note: We did not see the opposite usage where the note-taker would ask “Do you mean the French poet Villon?” although it is obviously possible.

R.3 Correction specified by context (19). The note-taker asks based on understanding the context (example: D: “We meet at 5 at the church.” N: “am or pm?” D: “afternoon, today afternoon.”).

R.4 Correction specified by order (6). The note-taker asks for a specific part of dictation using item order (example: D: “Buy oranges, apples and butter.” N: “What was the second one?”).

R.5 Correction by repeating a preceding text segment (36). User repeats the whole phrase up to the moment where s/he needs re-dictation (example: N: “Drive towards..?” D: “toward the castle”).

R.6 Correction by repeating a shorter form (1). This is similar to the above repeating of a preceding text segment, but in this case the dictating party only provides a shorter form of the corrected or missing segment, and it is up to the note-taker to finalize the text. Example: D: “Put it to microwave and warm it.” N: (only understood up to “microwave”) “microwave?” D: “warm it”. N: (expands to “and warm it”).

R.7 Spelling (3). Initially, the drivers tended to repeat a problematic phrase two or three times and if this was not understood, they started to spell that phrase. This strategy was used both by the driver and note-taker.

5.5 General Observations

G.1 Shortening dictated segments (2). The participants tended to utter shorter phrases in the noisy environment that what would be natural under silence.

G.2 Not speaking over high noise (N/A). Participants did not like to speak over high noise. Instead they waited or repeated the same segment (example: D: “Please, we will go, we will go, damn some motorcycle is starting here.”).

G.3 Multiple repeats result in complex acknowledgements (4). When multiple repeats are needed, participants tend to fall to more complex acknowledgement schemas (e.g. the above-mentioned acknowledgements by repeating).

G.4 Careful repeating (15). When repeating a phrase, the driver’s pronunciation is more careful, s/he tends to over-articulate.

Note: This is often an issue for man-machine systems as it poses a challenge for ASR models.

G.5 Noise adaptation (N/A). Users adapt to the noise level. What is totally unacceptable at the beginning becomes relatively fine after a period of time.

Note: This is observable even in the evolution of driving distraction statistics as reported in Section 6.1.

G.6 Resignation (4). After some period of repeating, the drivers sometimes resigned and agreed to an approximate version or even to a wrong meaning and proceeded further in dictation.

Note: We did not see the opposite pattern where the note-taker would resign although it is obviously possible.

G.7 Well known topics are easier to handle in noisy conditions (N/A). This holds also for local names, if they are unknown to the user, spelling has to be requested.

6 Discussion

The observations listed above appeared in our experiments with varying frequencies reported below in Fig. 2.

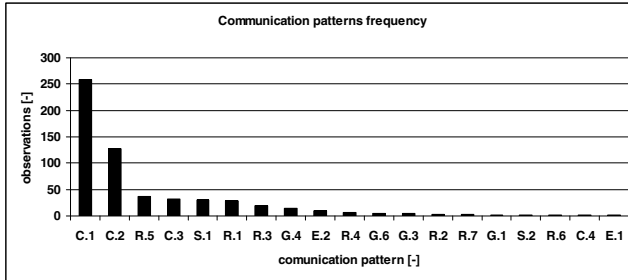


Fig. 2. Frequency of observations in recorded experimental data; see Section 5 for details

6.1 Impact of Dictation on Driver’s Attention

Fig. 3 summarizes distraction caused by the secondary task. Average mean deviation (MDEV) values with standard errors are depicted separately for each task. We adapted the ideal LCT track using the “Undistracted track 2”. The MDEV values for distracted driving were then compared against the “Undistracted track 1” and the difference was analyzed for significance. Participant 1 was excluded from the overall evaluation as an obvious outlier.

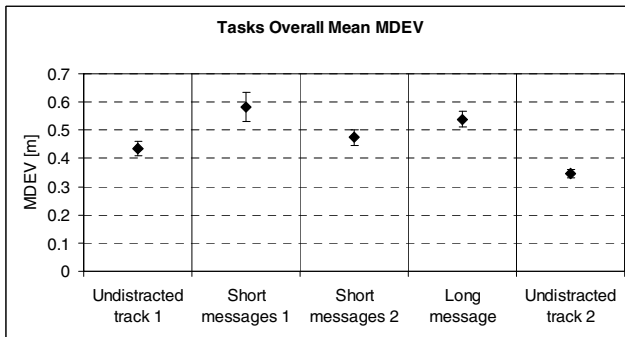


Fig. 3. Average overall mean deviation using data from all participants, except for participant 1 (outlier). Standard error of measurement is displayed.

Fig. 3 shows that driving while dictating under noise was more demanding from the driver’s point of view than undistracted driving (both undistracted tracks). The conventional two-tailed paired T-test for sample means confirmed statistical significance of the difference between mean values of MDEV for “Short messages 1” and “Undistracted track 1” ($p = 0.019$) and between “Long message” and “Undistracted track 2” ($p = 0.013$).

There is a visible difference between the tasks “Short message 1” and “Short message 2” which is caused by a learning effect (see Fig. 3). The difference between “Undistracted track 1” and “Undistracted track 2” is not only due to learning effect as “Undistracted track 2” was used for adaptation of the ideal track.

7 Conclusion

We presented here the results of collecting and analyzing human behavior when dictating both shorter and longer segments of text. Our study was conducted on 10 users, each of which drove 2 undistracted and 3 distracted LCT trips while dictating under noise. In the resulting transcribed data, we collected and described 22 elementary communication patterns and quantified their observed frequencies. The purpose of our work was to gain better understanding of natural usage patterns used by humans in order to utilize them for development of more natural human-computer dictation UIs and dialog systems in general.

As further steps in analyzing human to human communication patterns, we foresee several activities worth pursuing. Firstly the reported tests did not cover some types of distraction including echo or communication channel deterioration with dropping some words. Another interesting area is to proceed by analyzing behavior of the users with visual contact.

References

1. Brostrom, R., Bengtsson, P., Axelsson, J.: Correlation between safety assessments in the driver-car interaction design process. *Applied Ergonomics* 42(4), 575–582 (2011)
2. Cuřín, J., Labský, M., Macek, T., Kleindienst, J., Young, H., Thyme-Gobbel, A., Quast, H., Koenig, L.: Dictating and editing short texts while driving: Distraction and task completion. In: *Proceedings of the AutomotiveUI Conference*. ACM, New York (2011)
3. Mattes, S.: The Lane-Change-Task as a Tool for Driver Distraction Evaluation. In: *Proc. Annual Spring Conference of the GFA/ISOES* (2003)
4. Karat, J., Horn, H., Karat, C.: Overcoming unusability: Developing efficient strategies in speech recognition systems. In: *Proceedings of CHI 2000 Conference*, pp. 141–142. ACM, New York (2000)
5. Kun, A.L., Schmidt, A., Dey, A., Boll, S.: Automotive user interfaces and interactive applications in the car. In: *Personal and Ubiquitous Computing*, pp. 1–2 (2012)
6. Hart, S.G., Stayeland, L.E.: Development of NASA-TLX (task load index): Results of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (eds.) *Human Mental Workload*. Amsterdam North Holland Press (1988)
7. Pauzie, A.: Evaluation of Driver Mental Workload Facing New In-vehicle Information and Communication technology. In: *IET Intelligent Transport Systems, Special Issue – Selected Papers from HCD* (2008)
8. Yun-Cheng, J., Paek, T.: A Voice Search Approach to Replying to SMS Messages. In: *Proc: INTERSPEECH 2009, 10th Annual Conference of the Intl. Speech Communication Association*, Brighton, United Kingdom (2009)