

Experimental Validation of TranScribe Prototype Design

Hao Shi

Velian Pandeliev

ABSTRACT

In this paper we describe an experiment to compare a new transcription design prototype to an existing commercial application. In a between-subjects design, participants transcribed short interview segments using both applications, and they were compared with regards to quantitative efficiency measures such as time taken, and also on subjective user ratings assessed using Likert scales on an exit questionnaire. Discouragingly, most measures found no significant improvements of our prototype on the commercial product, although user satisfaction was significantly higher in the prototype. Over longer transcription periods and with additional user-supporting features, we expect our application to yield greater efficiency and even higher satisfaction, although some design features such as shortcut key placement will have to be improved.

INTRODUCTION

We have been interested in developing a software transcription tool to help occasional transcribers to produce accurate transcriptions efficiently. A heuristic evaluation gave us the confidence that we had identified the major issues with our design, and we created a functional prototype in C#, which would allow us to evaluate our design not only through expert reviews, but in a rigorous quantitative study. We traced our improvement decisions back to the beginning of the process, when we gathered user requirements on a commercial transcription tool called Express Scribe. The problems identified in that requirements study served to guide us in designing TranScribe, and so the next step in the design process would intuitively be to use Express Scribe as the benchmark for our prototype in terms of efficiency, user satisfaction and on the principal design questions we have been faced with.

The study described below was designed as a direct comparison between the two transcription applications, which has the obvious drawback of comparing a fully tested commercial application to a functional prototype coded over the course of a week. Nevertheless, the prototype of TranScribe (see Appendix A) implemented the same audio playback and text editing features that were used in Express Scribe, and was created to minimize confounds in the experiment.

METHOD

Participants

Six adults (2 female, 4 male) with normal or corrected-to-normal visual acuity were recruited for the experiment.

Design

In this experiment, our prototype was compared directly to Express Scribe, the commercial transcription application that was evaluated for usability issues at the beginning of this project. In a repeated-measures design, each participant was asked to transcribe two one-minute segments of the same interview, once using Express Scribe, and once using TranScribe. The two segments appeared in the same order, but the transcription applications were counterbalanced to minimise primacy and fatigue effects. Each transcription was timed as a measure of efficiency. In our previous experiments, it was determined that users resort to the mouse when they cannot find a way to accomplish a task efficiently with the keyboard, which is why the number of mouse clicks was recorded for each condition as an indication of interrupted typing flow. Then, a 7-point Likert scale questionnaire obtained subjective measures of the ease and enjoyment of using each application, their hotkey placements, and the participant's overall preference between the two (see Appendix A).

Since TranScribe was created specifically to address problems encountered with Express Scribe, it was hypothesized that transcriptions done in TranScribe would be faster than those done in Express Scribe, that TranScribe would score higher in overall usability and shortcut key placement, and that it would be preferred by participants overall.

Apparatus

Experiments were conducted on a 15.4" widescreen portable computer with a screen resolution of 1280x800 using an external keyboard and mouse. Green painter's tape was applied to the function keys which served as shortcuts to Express Scribe and TranScribe and they were labeled appropriately. Participants were fitted with a headset and the volume level was adjusted individually during the practice transcription. Both applications allowed for audio playback at two different speeds (50% and 100%) while preserving pitch, and had fine-grained rewind and fast forward functionality. They also had different shortcut key maps (See Fig. 1) Transcriptions were timed to the nearest second with the stopwatch function on a portable media player and the number of mouse clicks were counted by one of the experimenters.

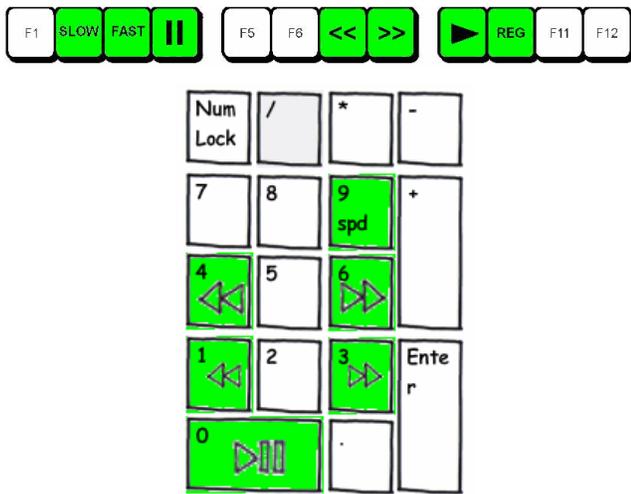


Figure 1. Express Scribe (above) and TranScribe (below) playback shortcuts.

Procedure

Participants were seated in front of the experimental PC in a quiet conference room. They were given an instructions form and asked to sign an informed consent (see Appendix B). An experimenter described the steps to the experiment, and they were introduced to the first transcription interface. They were allowed to familiarize themselves with the playback and the flow of the interface on a practice audio file from a different interview. Then they completed the transcription. They were asked to identify speakers by A and B, and to produce a transcription of good quality efficiently. The same steps were repeated for the other transcription application.

RESULTS

Efficiency

The mean time to complete a transcription in Express Scribe was 8 minutes and 30 seconds, and in TranScribe it was 8 minutes and 52 seconds. The difference was found to be non-significant using the t-test statistic (paired $t(10) = 0.137$, $p > 0.894$). Mouse click analysis was complicated by very sparse use of the mouse (less than 3 clicks per condition in general) and one big outlier. Nevertheless, means were analysed and the difference in mouse clicks was found to be non-significant (paired $t(10) = 1.025$, $p = 0.330$).

Subjective Data

The following table shows the median Likert scale values for each condition.

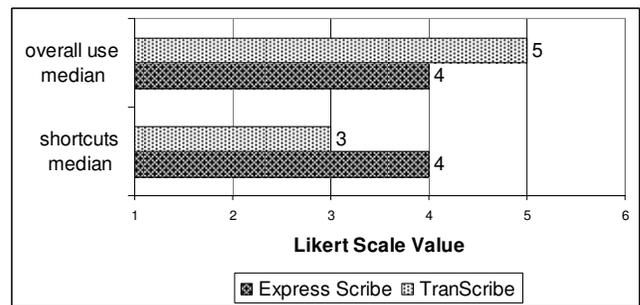


Table 1. Median Likert scale (1-7) values of overall use and shortcut satisfaction.

The next table shows the mean Likert scale values for each condition.

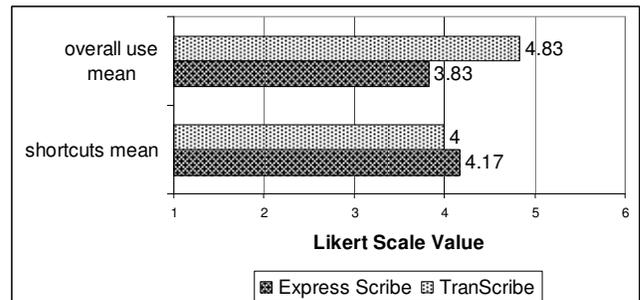


Table 2. Median Likert scale (1-7) values of overall use and shortcut satisfaction.

The differences were analysed using the t-test statistic. While the shortcut preference was not found to be significant (paired $t(10) = 0.170$, $p = 0.868$), participants did express a significant preference for TranScribe ($m_{ts} = 4.833$) over Express Scribe ($m_{es} = 3.833$) overall (paired $t(10) = 2.301$, $p = 0.044$).

The final question on the questionnaire was to express a preference for one or the other directly. It was on a modified Likert scale (see Appendix A) to eliminate bias towards TranScribe. The mean and median preference reported was +0.5 towards TranScribe.

DISCUSSION

At the beginning of this project we conducted a usability analysis of Express Scribe which included measures of the amount of time participants took to transcribe a minute of an audio interview. While no direct comparison can be made to those ratios (partly because the interviews chosen in the first experiment were heavily British-accented, which would impact the overall time), we set out to create an application that would improve on the efficiency of Express Scribe through a more intuitive interface and better shortcut keys, which were the principal subjective complaints with the Express Scribe design. For this reason, additional features such as timestamp insertion, bookmarks and speaker identification were not tested in the functional prototype, even though they would impact the overall transcription time in a situation when they would be necessary. We wanted to focus on discovering whether

TranScribe offered a significant efficiency improvement to Express Scribe in the areas of general interface design and shortcut key placement.

Our chief measure of efficiency was time. As reported, TranScribe did not significantly improve transcription time. Moreover, time means for TranScribe were in fact *worse* than for Express Scribe. This result is discouraging, and we must look at other factors to identify the reasons for it.

Mouse clicks were initially believed to be an appropriate measure of interrupted keyboard flow, although the observed use of the mouse diminished the experimenters' confidence in that measure. Its lack of significance could be thought of as consistent with the main efficiency measure, or it could be a result of a poorly framed measurement.

A much better indicator of interrupted keyboard flow, something we had hoped to minimise, was the reported usefulness of our shortcut map. In the usability assessment of Express Scribe, our evaluators reported issues with the shortcut keys which included separate play and pause buttons, counterintuitive placement, lack of mapping between the on-screen button order and the shortcut key order, and finger travel distance to the Function key row. Our design moved the shortcut keys to the Number pad to address the issue, and provided a closer mapping between the keyboard and screen (see Figure 2). However, participants reported that the finger travel time to the number pad was longer and more inconvenient than to the Function key row. Observations also indicated that participants moved their hands only as far as they had to, often pressing the number pad buttons with their pinky fingers, which are likely to be less frequently used and have poorer dexterity. The subjective satisfaction difference between the two shortcut maps was not significant, but it did also favour the Express Scribe placement, which leads us to conclude that, at best, our shortcut key placement is not significantly better than Express Scribe's, and that we should use the Function key row in future iterations.

However, when participants assessed the overall ease of use and enjoyment of the two products, they were significantly more satisfied with TranScribe's design. They also expressed a non-significant, yet encouraging slight preference towards using our software despite the shortcut issues. We take this as an indication that TranScribe, even in its prototypical form devoid of additional labour-saving functions is an improvement upon Express Scribe, at least in user satisfaction and enjoyment. It is possible that these effects would be amplified over longer periods of use. If the iteration timelines allowed, we would have loved to run another experiment, comparing longer use of Express Scribe to a more full-featured version of TranScribe which also used the Function row keys.

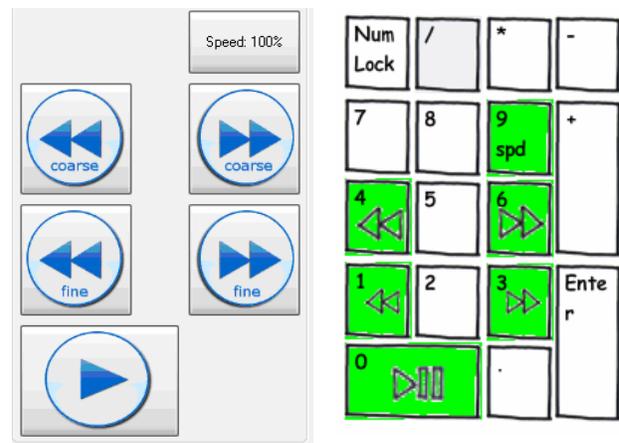


Figure 2. TranScribe on-screen and keyboard audio controls.

It is also worth noting that, while the chi-square statistic is recommended over the t-test for use with Likert scales, it was deemed unsuitable for our data because it requires a minimum of five instances of each value to be properly applied [3].

CONCLUSION

While our experiment did not elevate our product to the status of the next revolutionary step in audio transcription, it did serve to validate several design decisions, and to identify design problems such as our shortcut key placement. It is curious that this issue was mentioned in heuristic evaluation, but was deemed low severity by our evaluators. This demonstrates that there are design issues which require a quantitative experiment to be properly diagnosed and assessed for severity. We are nevertheless concerned with the lack of significant improvement in most of our measures, and we are working on improving the user's efficiency to match their level of satisfaction with our product.

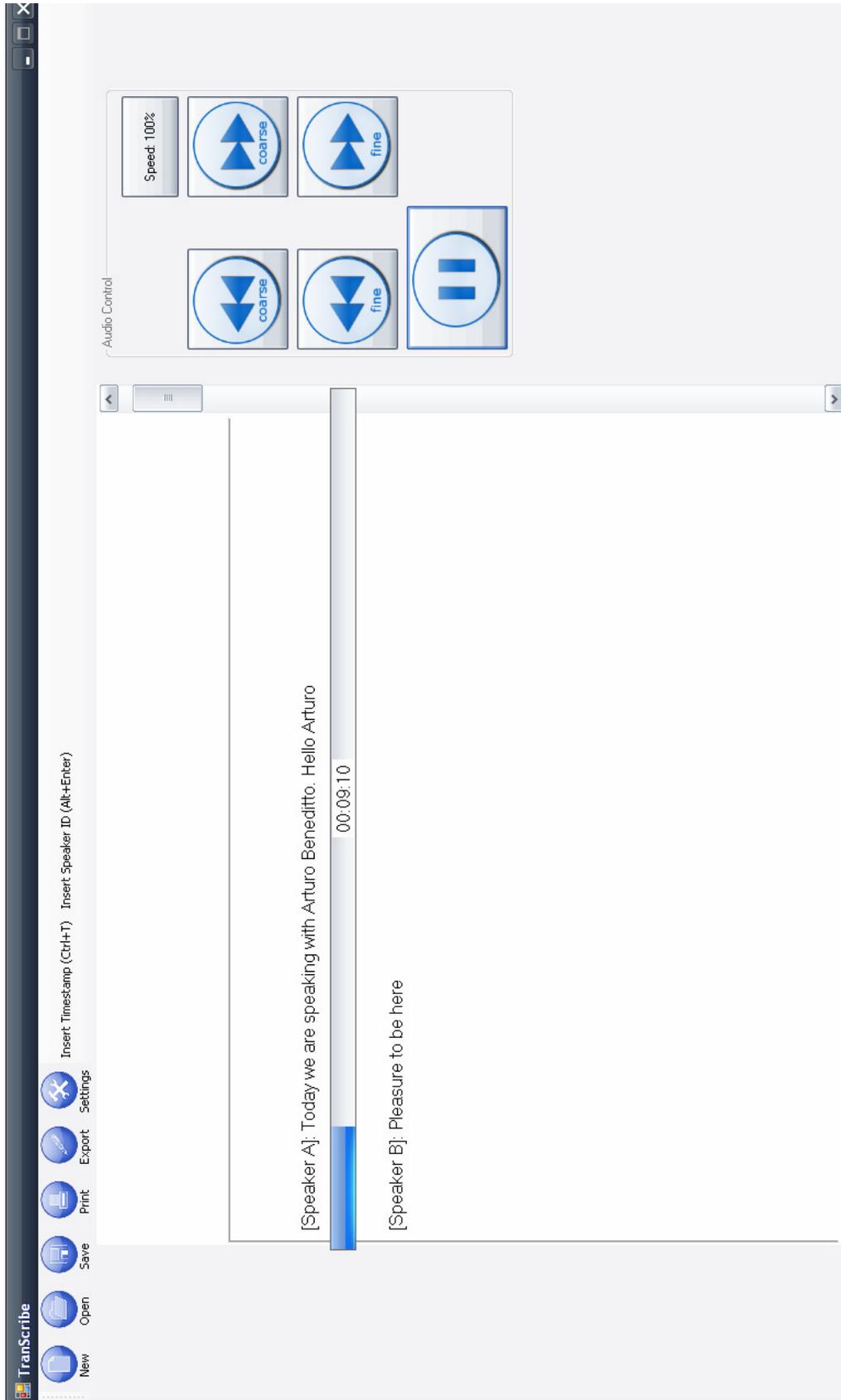
ACKNOWLEDGMENTS

We thank the six participants who volunteered for our experiment.

REFERENCES

1. Balsamiq Mockup Tool
<http://www.balsamiq.com/>
2. Gravetter, F. J. & Wallnau, L.B. Essentials of Statistics for the Behavioral Sciences. 5th ed. Wadsworth Publishing, 2004
3. SPSS Techniques Series: Statistics on Likert Scale surveys. University of North Iowa
<http://www.uni.edu/its/us/document/stats/spss2.html>
4. Pandeliev, V. & Shi, H. User Requirements for Audio Transcription. CSC2514 Assignment 1.

Appendix A: Prototype Screenshot



Appendix B: Instruction, Informed Consent and Questionnaire

Instructions:

This study aims to determine the efficiency of a new user-improved transcription design. You will be asked to transcribe two one-minute sections from an interview, one using a commercially available transcription application and one using a newly designed prototype. Please try to be as efficient as possible for both transcriptions. Before each segment, a short practice session will familiarize you with each transcription product. Keyboard shortcuts are indicated in green on the keyboard.

This study poses no significant physical, emotional or health risks to participants. Any data collected will be coded to preserve your anonymity and you are free to withdraw at any point without penalty.

Please sign below to indicate your consent for participating in this study.

Participant: _____

Witness: _____

Date: _____

Participant Number: _____

ES

Order	
Time	
Clicks	

What did you think of using Express Scribe for transcription?

1 2 3 4 5 6 7
Horrible Decent Excellent

How did you find the Express Scribe hotkey placement?

1 2 3 4 5 6 7
Horrible Decent Excellent

TS

Order	
Time	
Clicks	

What did you think of using TranScribe for transcription?

1 2 3 4 5 6 7
Horrible Decent Excellent

How did you find the TranScribe hotkey placement?

1 2 3 4 5 6 7
Horrible Decent Excellent

Which transcription software would you prefer to use?

4 3 2 1 2 3 4
Definitely ES Either Definitely TS