User Preference for Turntaking Tones 2: Participant Source Issues and Additional Data

TR 29.3447

October 6, 2001

Melanie D. Polkosky
James R. Lewis

IBM Voice Systems

West Palm Beach, Florida

Abstract
In this experiment, 24 participants indicated their preference for four tones as potential indicators of a user's turn to speak in speech recognition applications. The participants came from two different sources: IBM employees and externally-hired temporary employees. They rated four tones (presented using a Greco-Latin design) generated from synthesized sounds (flute, oboe, castanets, sine wave) for both pleasantness and appropriateness. Results indicated both participant groups rated the tones similarly and preferred the flute tone. The report concludes with a discussion of the implications of these results for experimental design and auditory interface design.

ITIRC Keywords

Turntaking
Speech recognition application
Audio tones
Auditory interface design
User preference

Contents

Introduction

A previous study (Polkosky, 2001) investigated listener preferences for four tones as potential turntaking cues for speech recognition applications. The participant sample included 16 individuals hired from a temporary agency (balanced for gender and age). The results indicated that the listeners rated a flute tone as both more pleasant and appropriate than tones derived from an oboe, sine wave, or castanets.

Although it may be desirable to recruit participant samples that are representative of an external population, time and financial constraints can make such samples difficult or impractical to obtain. In the current study, I replicated the design of Polkosky (2001) to determine whether there were preference differences between IBM employees and external participants. This study is of interest because it can help establish whether systematic differences occur in subjective ratings of auditory tones based on the source of participants.

Method

Participants. Twenty-four individuals completed this study. They comprised two participant groups: a non-IBM group of 16 individuals hired from a temporary employment agency for the purpose of participating in this and other unrelated human factors studies (Polkosky, 2001) and an IBM group of eight employees from the West Palm Beach, Florida location. Both groups were balanced for gender and age (above and below 40 years).

Stimuli. Participants heard four tones, counterbalanced to reduce order effects across participants. I used a Yamaha PSR-540 XG MIDI portable keyboard[1] synthesizer to generate three instrumental tones. The fourth tone (sine wave) was generated in Sound Forge 4.5d[2] and is the turntaking tone currently used in speech recognition applications by IBM Voice Systems. All tones had 16 bit sample sizes and 44,100 Hz sample rates (edited in Sound Forge 4.5d). Specifically, the four tones were:

A. Oboe: This tone's source was the steady state of a C4 note played by an oboe voice. Its prominent frequency occurs at 431 Hz (note A4) and it has a duration of 84 ms.
B. Castanets: This tone's source was the steady state a single C4 note played by a castanets voice. Its prominent frequency occurs at 7738 Hz (note B8) and it has a duration of 50 ms.
C. Sine wave: The prominent frequency of this tone occurs at 410 Hz (note G#4) and it has a duration of 69 ms.
D. Flute: This tone's source was the steady state of a C4 note played by a flute voice. Its prominent frequency occurs at 1056 Hz and it has a duration of 61 ms.

I selected these tones from 11 potential tones based on review by an expert panel (IBM human factors engineers) and recent literature, which suggests that tones should be "sinusoidal" and medium to high pitch (Balentine & Morgan, 1999). In addition, several studies suggest that musical tones are appropriate for use as auditory cues in human-computer interaction (Blattner, Sumikawa & Greenberg, 1989; Brewster & Crease, 1999; Brewster, 1998; Gaver, 1997) and users rate them as more pleasant than environmental sounds (Sikora, Roberts, & Murray, 1995).

To simulate use of the tones in a speech application, each tone occurred at the end of a spoken prompt for a news information service. The spoken prompt was a statement with falling intonation ("Sports news or results"), spoken by a female and recorded into Sound Forge 4.5d (16 bit, 44,100 Hz). This prompt helped to prevent potential participant ratings based on a direct question (e.g., "Do you want sports news or results?"), which may elicit a user response (Brehms, Rabin, & Waggett, 1995).

[1] Yamaha is a registered trademark of the Yamaha Corporation.
[2] Sound Forge is a trademark of Sonic Foundry, Inc.

Procedure. The study used a digram-balanced Latin square design to counterbalance the order of presentation of the tones[3]. Table 1 shows the experimental design (run once for the IBM group and twice for the group of external participants)[4].

Table 1. Experimental Design

Participant Group	Order			
Non-IBM	A	D	B	C
Non-IBM	B	A	C	D
Non-IBM	C	B	D	A
Non-IBM	D	C	A	B
Non-IBM	A	B	D	C
Non-IBM	B	C	A	D
Non-IBM	C	D	B	A
Non-IBM	D	A	C	B

Table Abbreviations: A=oboe, B=castanets, C=sine wave, D=flute

Each participant read a brief description of the task, explanation of a turntaking tone, and two questions eliciting ratings (see Appendix A for participant instructions). Each participant received verbal clarification and additional explanation as needed. They then listened to a prompt/tone combination (played over desktop stereo speakers attached to a PC-compatible computer) and rated the tone's pleasantness and appropriateness using two bipolar 7-point rating scales (shown in Appendix A) (Fusco & Katz, 1992; Sikora, Roberts, & Murray, 1995). Participants repeated this procedure for the remaining three tones.

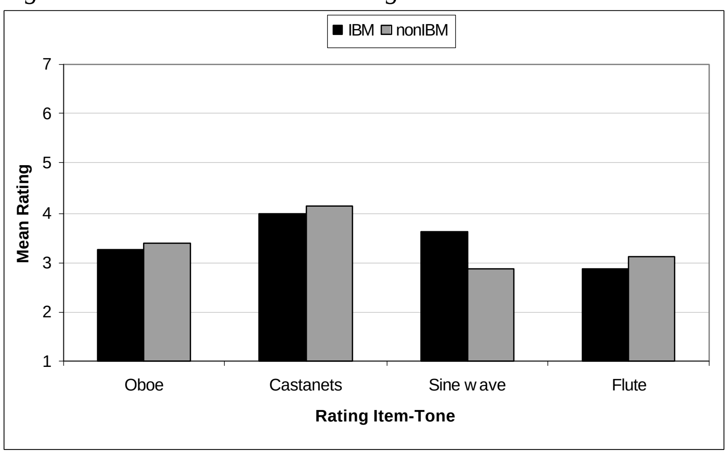
[3] This scheme not only results in standard Latin square counterbalancing of order of appearance in rows of the design, but also controls immediate sequential effects (Bradley, 1972; Lewis, 1993).
[4] Note that the data for the external participants in this report is the same as the data reported in Polkosky (2001). The data for the IBM employees is new, allowing comparison of results from the two participant groups.

Results

A mixed model ANOVA indicated only a significant main effect of tone ($F(3,66)=7.74$, $MSe=2.144$, $p<0.0001$). The main effect of participant group was not significant -- in fact, strikingly nonsignificant ($F(1,22)=0.002$ $MSe=14.772$, $p=0.969$), indicating no difference in the two groups' ratings. No other main effects or interactions were significant ($p>0.14$).

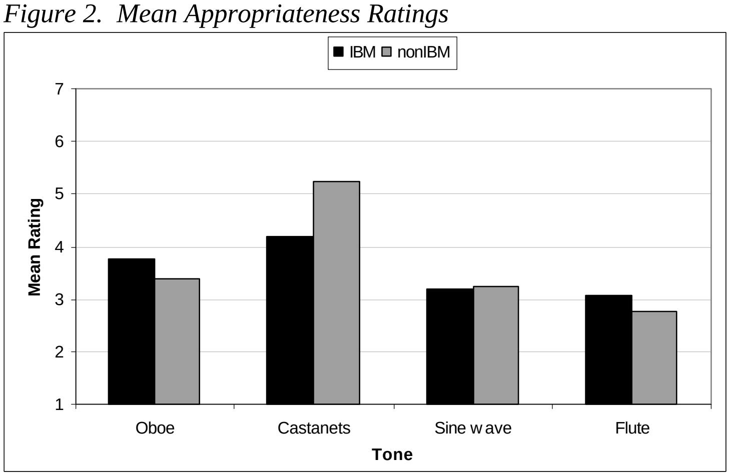
Pleasantness. Figure 1 presents the mean subjective ratings for pleasantness (lower scores are more pleasant). Overall, participants rated the flute tone as most pleasant (*m*=2.96, *sd*=1.78), followed by the oboe (*m*=3.29, *sd*=1.88), sine wave (*m*=3.38, *sd*=1.71), and castanets (*m*=4.04, *sd*=2.18). These results indicate a similar pattern for the IBM employees and the non-IBM data previously documented in Polkosky (2001).

Figure 1. Mean Pleasantness Ratings



Appropriateness. Figure 2 presents the mean subjective ratings for appropriateness (lower scores are more appropriate). Participants rated the flute tone as most appropriate (*m*=2.96, *sd*=1.78), followed by the sine wave (*m*=3.21, *sd*=1.96), and oboe tone (*m*=3.63, *sd*=1.88). They rated the castanets tone as least appropriate (*m*=4.54, *sd*=2.11). These results also indicate a similar pattern for the IBM employees to the data previously reported for non-IBM employees (Polkosky, 2001).

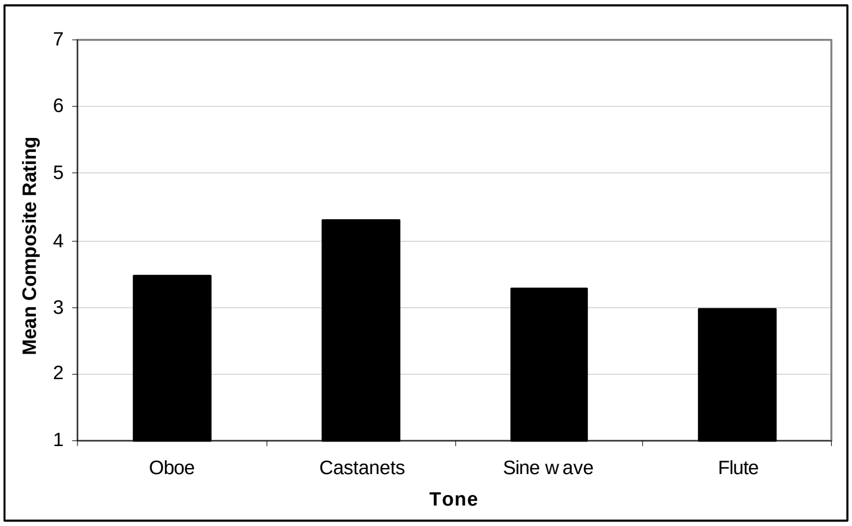
Figure 2. Mean Appropriateness Ratings



Combined analysis. Because the main effects of rating and participant source were not significant in the ANOVA ($F(1,22)=0.265$, $p=0.612$), additional analyses used a composite score (the mean of pleasantness and appropriateness ratings across both participant groups).

Figure 3 shows the mean composite score for each tone. Post hoc *t*-tests indicated the following mean differences were significant with the Bonferroni correction (*alpha*=0.02): oboe-castanets ($t(23)=-2.83$, $p=0.009$), castanets-sine wave ($t(23)=2.70$, $p=0.01$), and castanets-flute ($t(23)=3.79$, $p=0.0009$). The castanets tone was clearly inferior to the other three tones. The mean difference between the oboe and flute tones approached significance ($t(23)=2.24$, $p=0.035$). All other comparisons were not significant ($p>0.2$).

Figure 3. Mean Composite Rating by Tone



Discussion

This study is a replication of Polkosky (2001) and provides additional data regarding preferences for IBM employees to the data already existing for external participants. In the study, two participant groups rated their preference for four tones as potential turntaking cues in speech recognition/telephony applications. Two participant groups (one recruited from IBM employees, the other hired from a temporary agency) rated the tones similarly. Additionally, the flute and sine wave tones received the highest composite preference ratings, suggesting users prefer these tones as a potential turntaking cues in speech recognition/telephony applications. Participants rated the castanets tone as least pleasant and appropriate.

These results are important because they provide some evidence that internally-recruited participant ratings may be generalized to a broader population of non-IBM employees. Although other studies of auditory tones suggest gender and age differences in tone preferences (Polkosky & Lewis, 2001), this study provides some preliminary evidence that the source of participants does not necessarily influence results. If similar results occur with other auditory cues (e.g., waiting or landmarking tones), they would suggest that tone preference studies may use internal employees (balanced for gender and age) without sacrificing generalizability. Recruiting internal volunteers allows such studies to be relatively economical, rapidly executed, flexible, and simple, especially when they support ongoing development efforts.

The current results also provide additional support for the findings of Polkosky (2001). In that study, I found initial indication that users prefer a flute tone. The current results show users do prefer the flute tone but consider it similar to the sine wave tone. Therefore, designers may confidently use the flute or sine wave tone as a turntaking cue in speech recognition/telephony applications.

References

Blattner, M., Sumikawa, D., & Greenberg, R. (1989). Earcons and icons: Their structure and common design principles. *Human-Computer Interaction, 4,* 11-44.

Bradley, J. V. (1958). Complete counterbalancing of immediate sequential effects in a Latin square design. *Journal of the American Statistical Association*, *53,* 525-528.

Brems, D., Rabin, M., & Waggett, J. (1995). Using natural language conventions in the user interface design of automatic speech recognition systems. *Human Factors, 37(2),* 265-282.

Brewster, S. (1998). Using nonspeech sounds to provide navigation cues. *ACM Transactions on Human Computer Interactions, 5(2),* 224-259.

Brewster, S. & Crease, M. (1999). Correcting menu usability problems with sound. *Behavior and Information Technology, 18(3),* 165-177.

Fusco, M. & Katz, R. (1992). Catch a rising tone: Selecting a tone for a new calling service. In Proceedings of the Human Factors Society 36th Annual Meeting, (pp. 227-231). Santa Monica, CA: Human Factors and Ergonomics Society.

Gaver, W. (1997). Auditory interfaces. In M. Helander, T. Landauer, P. Prabhu (Eds.), *Handbook of Human-Computer Interaction*, 2nd ed. (pp. 1003-1041). Amsterdam: Elsevier.

Lewis, J. R. (1993). Pairs of Latin squares that produce digram-balanced Greco-Latin designs: A BASIC program. *Behavior Research Methods, Instruments, & Computers*, *25*, 414-415.

Polkosky, M. (2001). User preference for turntaking tones. (Tech. Report No. 29.3420). Raleigh, NC: International Business Machines.

Polkosky, M. & Lewis, J. (2001). Effect of ticking rate on user estimates of system response time. (Tech. Report No. 29.3425). Raleigh, NC: International Business Machines.

Sikora, C., Roberts, L., & Murray, L. (1995). Musical vs. real world feedback signals. In *Proceedings of CHI'95*. (pp. 220-221). New York: ACM.

Appendix A. Instructions to Participants

You will hear a recorded voice, followed by a short sound. This sound may be used as a special type of sound called a TURNTAKING TONE. The turntaking tone is used when a person is talking with a computer. The turntaking tone tells you that it is your turn to speak to the computer.

Please indicate whether the sound is PLEASANT and APPROPRIATE on the scale shown below (circle appropriate number). Please feel free to add comments about the sound in the space provided or on the back of this sheet.

1. How **pleasant** is this sound?

1	2	3	4	5	6	7
Very Pleasant						Very Unpleasant

Comments:

2. How **appropriate** is this sound to tell you it is your turn to speak to the computer?

1	2	3	4	5	6	7
Very Appropriate						Very Inappropriate

Comments:

Appendix B. Participant Data

Participant Group	Oboe Tone		Castanets Tone		Sine Wave		Flute Tone	
	Pleasant	Approp	Pleasant	Approp	Pleasant	Approp	Pleasant	Approp
nonIBM	3	3	6	6	2	2	2	2
nonIBM	6	6	6	4	5	6	6	5
nonIBM	6	6	5	5	5	6	6	6
nonIBM	2	3	2	6	4	4	2	2
nonIBM	3	5	6	6	4	6	5	6
nonIBM	1	1	1	1	1	1	1	1
nonIBM	1	1	2	1	2	1	2	2
nonIBM	7	1	7	1	5	1	3	1
nonIBM	4	3	6	2	7	1	3	1
nonIBM	1	6	1	5	2	5	2	5
nonIBM	1	1	1	1	1	1	1	1
nonIBM	3	5	7	5	6	5	2	5
nonIBM	1	7	1	7	1	1	1	3
nonIBM	2	3	3	5	3	4	2	2
nonIBM	5	5	5	5	5	5	2	2
nonIBM	6	4	5	7	5	2	6	5
IBM	3	4	6	6	3	4	5	4
IBM	3	4	3	5	2	2	2	3
IBM	4	2	7	7	2	2	1	1
IBM	3	4	5	7	3	4	5	4
IBM	5	6	3	3	4	7	5	5
IBM	3	1	3	3	5	3	2	1
IBM	5	4	5	7	2	2	4	3
IBM	1	2	1	4	2	2	1	1