

# Multimodal information Management: Evaluation of Auditory and Haptic Cues for NextGen Communication Displays

*Durand R. Begault (Code TH, NASA Ames Research Center)*

*Rachel M. Bittner (New York University)*

*Mark R. Anderson (Dell Systems)*

Auditory communication displays within the NextGen data link system may use multiple synthetic speech messages replacing traditional ATC and company communications. The design of an interface for selecting amongst multiple incoming messages can impact both performance (time to select, audit and release a message) and preference. Two design factors were evaluated: physical pressure-sensitive switches versus flat panel “virtual switches”, and the presence or absence of auditory feedback from switch contact. Performance with stimuli using physical switches was 1.2 s faster than virtual switches (2.0 s vs. 3.2 s); auditory feedback provided a 0.54 s performance advantage (2.33 s vs. 2.87 s). There was no interaction between these variables. Preference data were highly correlated with performance.

## INTRODUCTION

The increased operational autonomy of flight crews in the NextGen environment will potentially result in higher overall workload and greater demands on visual and auditory modalities to safely interact with automation and the overall complexity of the future flight environment. This increase in information transfer poses a scientific and engineering challenge to the way flight crews interact with The Federal Aviation Administration has recognized the challenges of “improving flight crew awareness” in NextGen (FAA, 2001):

With the increased role of automation, maintaining flight crew awareness and effective intervention during failure and abnormal conditions is critical...AVS must develop new displays and alerting, as appropriate, to improve awareness and retain the ability for the flight crew to manage the operation.

The National Research Council’s “Decadal Survey of Civil Aeronautics: Foundation for the Future” identified as a high priority “Interfaces that ensure effective information sharing and coordination among ground-based and airborne human and machine agents”, and to “Interfaces and procedures that support human operators in effective task and attention management” (NRC, 2006). To mitigate overloading of the visual perceptual system, auditory displays can be enhanced beyond normal radio communications and caution-warning signals to include synthetic speech messaging. These messages can convey data such as flight status and trajectory for shared situational awareness amongst aircraft and from aircraft to ground control.

One possible scenario for an improved auditory display within the NextGen data link system will involve multiple synthetic speech messages replacing traditional ATC and company communications advising weather, routing changes, and status of nearby aircraft (Rehman and Mogford, 1996). These may also be virtual versions of “party line information” where the communications from other surrounding aircraft may be audited for situational awareness (a form of audio “twittering”). Many of these messages will become more essential in air operations with increasing autonomous decision-making and need for situational awareness regarding nearby aircraft.

Multiple speech communication messages may be sent to the flight deck from automated systems, resulting in a *message array*, or a set of messages that cannot be audited in real time, much like reading an email inbox. Therefore, there will be a need for a system that organizes such messages such that they may be selected for listening in terms of priority or age of the message. Most of these will be non-critical and therefore may be reviewed by the pilots at preferred times. Message arrays are currently familiar in the flight deck from the stream of text messages that can be selected by scrolling and reading. However, there are temporal advantages to acquiring information via listening as opposed to reading, particularly when complex visual displays require constant vigilance.

Proposals have been made for improving NextGen auditory displays, such as implementation of spatial auditory cues, and varying the prosody and speaking rate of stored speech. By contrast, investigations of designs to accommodate the user’s interaction with auditory displays have been for the most part neglected. Although speech recognition is one means by which interaction might be accommodated, unacceptably high error rates and temporal lag will make manual interaction with controls the more likely means of interaction.

Recently, touchscreen-based tablet computer to substitute for the use of traditional paper flight charts on commercial airline flights (Bilton, 2012). Flat-panel touchscreens have obvious advantages over traditional manual controls and displays, including ease of reconfiguration and expense. Boeing cites advantages of flat screen displays in their background website on the 777 flight deck (Boeing, n.d.).

The depth of the new “flat panel displays” is about half that of CRTs. In addition to saving space, the new displays weigh less and require less power. They also generate less heat, which contributes to greater reliability and a longer service life. As another benefit, the displays do not require the heavy, complex air conditioning apparatus needed to cool equipment on previous flight decks. Pilots appreciate that flat panel displays remain clearly visible in all conditions, even direct sunlight.

Flat panel touch screens simulate actual buttons but are significantly different in that finger action is transmitted in terms of duration and area of contact by the finger, as opposed to a physical displacement caused by the finger. Another feature of traditional buttons and switches is that they provide auditory cues as to their status as a consequence of their mechanics, while touch panels must synthesize audio cues. The differences have potential implications both in terms of perceptual feedback, performance and for preference on the part of pilots to operate devices effectively.

## METHOD

Familiar multi-channel communication interfaces are represented by the design and physical layout of Bell Laboratories' multi-line telephone (Figure 1, top). Such telephones have been in wide use in the United States since the 1950s, and the user interface and performance characteristics have changed little between the classic Bell System-Western Electric telephones and their modern digital equivalents.

The multi-line telephone acts as the user's "interface" to the communication system, allowing the selection and organization of multiple incoming calls. The selection amongst various input lines in a chosen time sequence can be quickly accomplished by the user without regard to details of the communication system itself, such as the volume adjustment or frequency selection necessary on the flight deck radio. From a human factors standpoint, the multi-line telephone interface is elegant in its simplicity; its design conforms to the "good design" principles espoused by industrial designer Dieter Rams, in that the buttons make a product "useful", "understandable", "unobtrusive" and it involves "as little design as possible."

The buttons are mechanical two-state switches that provide several forms of haptic and visual cues as to their current status. In addition to haptic and visual cues, the telephone uses two audio alerts. Most often, the familiar classic telephone ring is used to direct visual attention to the telephone receiver so that the line associated with an incoming call can be answered. The audio alert is unspecific as to the incoming line (counting from left to right, buttons 2-5). However a second "buzzer" sound indicates that a call has come internally via the intercom, and is answered using the rightmost button (button 6). Otherwise, auditing the classic telephone ring requires a subsequent action for first visually identifying the appropriate flashing button (button 2, 3, 4 or 5), and then engaging it by pushing the button inward (a "visually-guided haptic target"). To handle multiple incoming calls by placing some on "hold" while answering others, the user operates two buttons in sequence in a conventional manner. If a second incoming call arrives while the first line is engaged, the user presses the leftmost, red colored "hold" button (button 1). A flashing light indicates an incoming call; a "winking" light indicates a held line.

We used this design as a guide for the features of a modern equivalent of a NextGen communication interface. Research conducted at our laboratory examined two interfaces for a prototype five-channel message storage system for on-demand audio data link message playback. The distinct tactile features of the two types of control systems were examined as

independent variables: a flat panel display (iPad) with virtual, graphic buttons that responded to time and placement of the finger (Figure 1, middle); and a "real button" interface using raised pressure sensors strain gauges) that responded to finger pressure (Figure 1, bottom).

Four channels were "normal" incoming messages on buttons 2-5, while one channel was reserved for "priority" messages, button 6. "Normal" messages were signaled by a brief buzzer alert sound, and "priority" messages were signaled by a bell sound. The "hold" button was also provided to allow users to pause listening to a normal message in favor of a priority message, and then return to it later. For each display, the presence or absence of an auditory "switch sound" based on the sound of an actual button engaging or disengaging was an additional independent variable. Performance was compared across conditions as a function of time to respond to an incoming message alert and the duration required for listening to a message. "Normal" and "priority" messages were evaluated separately. In addition, measures of subjective preference and subjective performance were gathered.



Figure 1. *Top*: Selection buttons used with Western Electric 564 "key line" to select and hold different telephone lines. In auditory display parlance, these can be considered "information streams"; as applied to NextGen, as an array of stored synthetic speech data link messages. *Middle*: touch pad version used in experiment; *Bottom*: pressure sensor version used to emulate physical switches. The sensors were raised from the surface and provided tactile feedback.

Table 1 summarizes the experimental design. Within subject comparisons were evaluated for two levels of button type ("real" and virtual) and two levels of auditory feedback from the buttons (none, or auditory cues for both touching and engaging the button). These were tests of the hypotheses that the presence or absence of auditory feedback would cause a significant difference in the dependent variables. Separate analyses were made for "normal" and "priority" messages.

MESSAGE TYPE		“NORMAL”	“PRIORITY”
REAL BUTTONS (STRAIN GAUGE)	NO AURAL FEEDBACK	10 SUBJECTS	10 SUBJECTS
	W/ AURAL FEEDBACK	32 TRIALS PER BLOCK	8 TRIALS PER BLOCK
VIRTUAL BUTTONS (IPAD TOUCHSCREEN)	NO AURAL FEEDBACK	FOUR CONDITION (3 BLOCKS PER CONDITION)	FOUR CONDITION (3 BLOCKS PER CONDITION)
	W/ AURAL FEEDBACK		

Table 1. Experimental design

Ten participants ran twelve continuous blocks lasting approximately 5 minutes each in a simplified flight simulation while operating the message storage system in response to a continuous flow of synthetic speech data link messages. They were instructed to accomplish the task of listening to messages and responding to them via prescribed procedures as quickly and as accurately as possible. In response to an aural alert cue, the subject acknowledged the incoming data link synthesized voice message (DLM) by selecting line 1-4 for audition, or line 5 if a “priority message”. Additional messages could arrive during playback of a selected message, giving the subject the option to put ongoing messages on hold to audition “newer” messages in whole or part. If a message arrived marked as priority, the subject was obligated to put other messages on hold to audition it.

The twelve blocks were randomized between the four conditions shown in Table 1: virtual button, no audio feedback; virtual button, audio feedback; real button, no audio feedback; real button, audio feedback. From each subject, there were a total of 96 responses per condition for normal messages and 24 responses per condition for priority messages. Successive message onset times were randomly varied to occur within an interval of 3 - 8 s. The duration of each message was approximately 5 s, in the form of <call sign> <flight number> <instruction>; for example “American 2 9 4, climb to 10,000 feet”.

These “Normal” messages were meant to emulate current communication frequency “party line” messages caused by ATC commands to the three aircraft nearest to ownship. Future NextGen systems may use synthesized speech messages in place of text data link, as a form of “spoken tweets”. “Priority” messages indicated communications for ownship.

The experiment was designed to motivate subjects to listen to the content of the message while performing various tasks with a computer screen and a mouse, thereby distracting them from attending solely to the communication device. To do this, a set of procedures was given during training blocks on how to respond to different messages, and then interact with a simplified “radar display”, as explained in Figure 2. Shortly after an incoming priority message, a question about its content was displayed on a second iPad and required a response within six s (five-alternative forced choice).

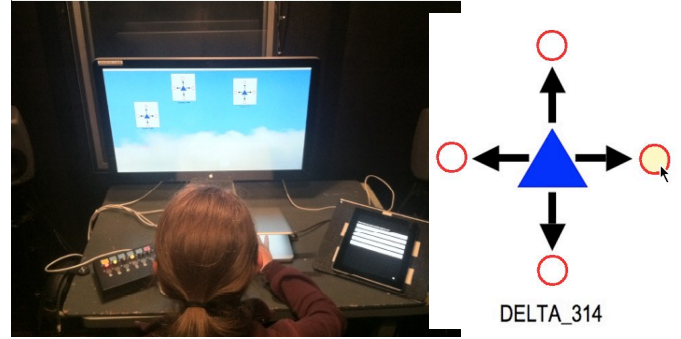


Figure 2. Left: Participant running an experimental block in booth. The “real” button interface is visible on the left; this device was replaced with an iPad for the touchpad experimental blocks. The tablet on the right presented a five-alternative forced choice question after a priority message was received (these data were not analyzed). Right: close-up of “radar” symbol seen on screen. Subjects were required to determine if the call sign indicated by the data link message corresponded one of the three available call sign identifiers, and then click the mouse within the circle corresponding to the instruction for that aircraft—climb, descend, turn left or turn right. The aircraft on the “radar” would change call sign and company intermittently to emulate the current proximate aircraft situation

The dependent variables were chosen to investigate interaction with the proposed interface in terms of performance and subjective evaluation. The dependent variables involving performance are shown in Figure 3, and include the time to respond to the aural DLM alert ( $t1$ ); time to complete audition of a single DLM ( $t2$ ); time in responding to priority messages ( $t3$ ); and frequency of use of the hold button. Subjective data were gathered at the conclusion of the experiment using a questionnaire, for comparative perceived self-performance under each of the four conditions, and separately for overall preference (hedonic rating). Data was gathered using both ratings (7 point scale) and rankings between the four conditions.

The dependent variable “message response time” was the time interval from the time an auditory cue for an incoming message began to the time the button was fully engaged to initiate playback of the message. Data for normal and priority messages were evaluated separately using a 3-way analysis of variance (ANOVA), with display type and presence of an auditory cue as the independent variables.

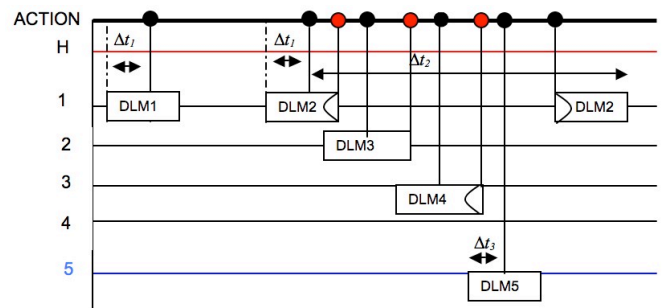


Figure 3. Sample data link message “score” for an experimental run. Time runs along the horizontal axes; the vertical axis indicates incoming message lines 1-4 and the priority line 5. DLM# = data link messages (looped). Dashed line = message alert cue (initiation). Black dot = activate line button; Red dot = activate hold button.  $\Delta t_1$  = message response time;  $\Delta t_2$  = message completion time..  $\Delta t_3$  = time to acquire a priority message (performance measure: message completion)

## RESULTS

Analysis of the raw timing data for “normal” messages indicated a significant effect for both audio cues ( $F = 5.574, p = 0.043$ ) and for interface type ( $F = 7.374, p = 0.024$ ). Stimuli using the “real” buttons were overall 1.52 s faster (2.15 s versus 3.67 s). Stimuli with supplemental audio cues were overall 0.52 s faster (2.65 s versus 3.17 s). There was no interaction between these variables. These data were analyzed in terms of the average value per subject and per condition. See Figure 4.

These stimuli included a number of outliers ( $> 3$  SD), which were removed for each condition in subsequent analyses. The percentage of outliers was 0.94-1.77% between the four conditions, out of the total 960 trials/condition. Analysis of the timing data for normal messages with outliers removed indicated a significant effect for both audio cues ( $F = 9.125, p = 0.014$ ) and for interface type ( $F = 12.68, p = 0.006$ ). Stimuli using the button box were overall 1.2 s faster than the touch pad (2.0 s versus 3.2 s). Stimuli with supplemental audio cues were overall 0.5 s faster than without supplemental audio cues (2.33 s versus 2.87 s). There was no interaction between these variables. See Figure 5.

Time response data are typically log transformed as  $\log_{10}(1+X)$  to normalize the distribution of response data (correct for positive skew). An ANOVA of the transformed data confirmed the significance results of the untransformed data analysis. Analysis of the log-transformed data indicated a significant effect for both audio cues ( $F = 6.429, p = 0.032$ ) and for interface type ( $F = 16.795, p = 0.003$ ).

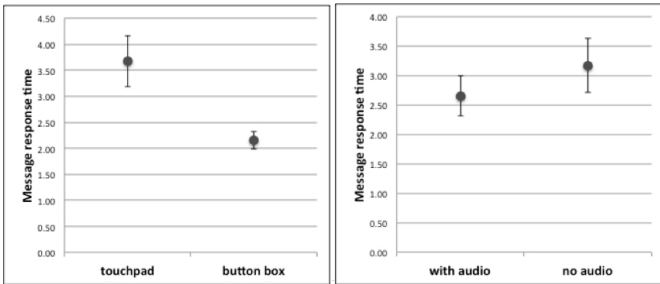


Figure 4. Raw Timing data, response time for “normal” messages. Significant main effects found for touchpad (virtual) versus button box (physical) switches, and for the presence of audio feedback made by activating the switch

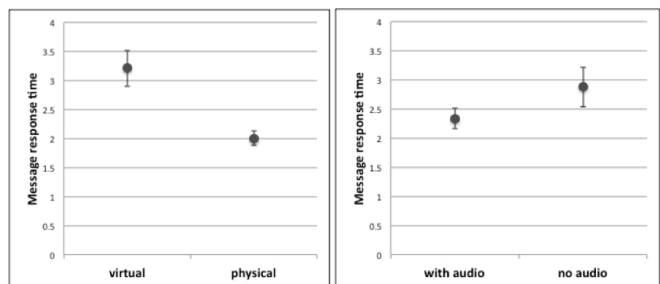


Figure 5. Timing data (with outliers removed) response time for “normal” messages. Significant main effects found for touchpad (virtual) versus button box (physical) switches, and for the presence of audio feedback made by activating the switch.

The raw values for priority message timing data indicated a total of 1.25-2.92% outliers ( $> 3$  standard deviations) within each of the four conditions, of the total of 240 trials per condition. Analysis of the timing data for priority messages with outliers removed indicated a significant effect for interface type ( $F = 58.164, p < 0.001$ ). Stimuli using the button box were overall 0.6 s faster (2.1 s versus 2.7 s). There was no significant effect of the audio cue or interaction. An ANOVA of the log-transformed data confirmed the significance results of the untransformed data analysis, with a significant effect for interface type ( $F = 29.213, p < .001$ ).

Message duration is defined as the average time duration measured from initiation to termination of a message from button activation. This is a measure of the ability for subjects to “clear” a completed message. A modest effect was noted for interface type, with the button box 0.3 s faster (4.9 s versus 5.2 s). There was no significant effect of the audio cue or interaction. The effect, while modest, could be additive and therefore more meaningful in the case of multiple messages.

The number of times a call was put on hold was on average 4 times per condition. There were no significant differences between conditions.

Participants provided ratings and rankings for perceived performance and preference for each of the four display conditions. The use of ratings and rankings may be considered redundant, but were included to insure consistency in responses. A Wilcoxon test was used to analyze these non-parametric data separately for each of the four rating evaluations across subjects. For the four treatments and 10 rank orders, the critical value is 14.8 for  $p \leq .05$ , and 18 for  $p \leq .01$  (Sachs, 1984). Table 2 below indicates a significant difference ( $p \leq .01$ ) between the virtual button without audio condition versus the “real” button with audio for perceived performance; Table 3 indicates similar results for preference. Other condition comparisons were only significant at  $p \leq 0.1$ .

PERCEIVED PERFORMANCE RATING

	IN	PN	IA	PA
	15	23	24.5	37.5
IN	15	8	9.5	<b>22.5</b> significant
PN	23		4	<b>14.5</b>
IA	24.5			<b>13</b>
PA	37.5			

PERCEIVED PERFORMANCE RANK

	IN	PN	IA	PA
	14	23	27	36
IN	14	9	13	<b>22</b> significant
PN	23	0	1.5	<b>13</b>
IA	27			9
PA	36			

Table 2. Wilcoxon pairwise comparison for questionnaire data regarding perceived performance (ratings were made on 7 point Likert scale). IN = virtual button without audio; PN = real button without audio; IA = virtual button with audio; PA = real button with audio.



PREFERENCE	RATING			
	IN	PN	IA	PA
	14.5	23	28	34.5
IN	14.5	8.5	13.5	20
PN	23		5	11.5
IA	28			6.5
PA	34.5			

PREFERENCE	RANKING			
	IN	PN	IA	PA
	15	24	25	36
IN	15	9	10	21
PN	24		1	12
IA	25			11
PA	36			

Table 3. Wilcoxon pairwise comparison for questionnaire data regarding preference (ratings were made on 7 point Likert scale).. IN = virtual button without audio; PN = real button without audio; IA= virtual button with audio; PA = real button with audio.

We also tested the association between our subjective and objective measures. A Spearman Rank Correlation coefficient was calculated by transforming the performance data into rank values (interval to ordinal conversion) and then comparing these ranks to the ranked opinion data.

Looking at each of the four conditions individually, the correlation between response time and preference rating was 0.72. The correlation between response time and perceived performance rating was 0.68. This indicates a moderately strong correlation between the quality metrics of perceived performance and preference, compared to the objective performance measure of response time. These values are near the 0.7  $p = .05$  significance level (using  $n-2$  degrees of freedom). Similar values were found for response time to priority messages ( $\rho = 0.62, 0.61$  for perceived preference and performance, respectively).

Additional analyses were conducted by averaging the response times for the two different audio conditions and then ranking according to interface condition, as well as by averaging the two different interface types and ranking according to audio condition. The justification is based on the significant effect found for the performance data for these main effects, and the lack of interaction. The correlation between response time and both preference and perceived performance ratings increases to 0.98 for normal messages when combining. For priority messages, the correlation between response time and both preference and perceived performance ratings increases to 0.96. The correlation between subjective and objective measures is therefore quite high when the main objective effect is accounted for.

## DISCUSSION

From the standpoint of interface design optimization, these results address NextGen prioritized concerns, as expressed in the National Research Council's "Decadal Survey of Civil Aeronautics" for "Interfaces that ensure effective information sharing and coordination among ground-based and airborne human and machine agents", and "Interfaces and procedures that support human operators in effective task and attention management". The results of the "trade study" aspect of this

investigation indicated that physical controls were significantly superior to virtual controls in terms of response time to an individual message by about 0.6-1.5 s. Audio feedback provided an advantage of about 0.5 s. When dealing with multiple messages in a real world context under high workload, these individual time advantages may combine to ensure safer, more efficient operations that foster effective information sharing and information management.

The subjective data results show a significant effect of preference and impression of superior performance for physical switches having audio feedback, compared to touch panel virtual switches. The correlation between objective measures of performance and subjective ratings of preference and performance was shown to be high.

Overall, the results indicate that any replacement of physical controls by virtual touch screens in NextGen flight deck controls must be considered carefully, and should include audio feedback. Additionally, the use of a five-channel message storage system with an on-demand playback interface shows promise for enabling pilots to successfully manage a complex set of NextGen data link messages.

## REFERENCES

- Bilton, N. (2011, December 14). F.A.A. Approves iPads in Cockpits, but Not for Passengers. *New York Times*. Retrieved from <http://bits.blogs.nytimes.com/2011/12/14/f-a-a-approves-ipads-in-cockpits-but-not-for-passengers/>
- Boeing (n.d.). Retrieved from [www.boeing.com/commercial/777/family/pf/background.html](http://www.boeing.com/commercial/777/family/pf/background.html)
- FAA (2011). Aviation Safety (AVS) Work Plan for NextGen 2011. Federal Aviation Administration. Retrieved from <http://www.faa.gov/nextgen/media/AVS%20Work%20Plan%20for%20NextGen%202011.pdf>
- NRC (2006). National Research Council Decadal Survey of Civil Aeronautics: Foundation for the Future. Washington, DC: National Academies Press. Retrieved from <http://www.nap.edu/>
- Rehmann, A. J., and Mogford, R. H. (1996). "Airborne data link study report." FAA technical report DOT/FAA/CT-TN95/62.
- Sachs, L. (1984). *Applied Statistics. A Handbook of Techniques*. Berlin: Springer-Verlag.