Synchronization of the acoustic evidence in the assassination of President Kennedy

R Linsker* and RL Garwin[1]
IBM T. J. Watson Research Center, P. O. Box 218 (1101 Kitchawan Road & Route 134), Yorktown Heights, NY 10598, USA
e-mail: linsker@us.ibm.com
H Chernoff[1]
Statistics Department, Harvard University, Cambridge, MA 02138, USA
P Horowitz[1] and NF Ramsey[1]
Physics Department, Harvard University, Cambridge, MA 02138, USA
Science & Justice 2005 45 207 – 226 Received 17 November 2004 accepted 23 May 2005

We have revisited the acoustic evidence in the Kennedy assassination – recordings of the two Dallas police radio channels upon which our original NRC report (Ramsey NF et al., Report of the Committee on Ballistic Acoustics. National Research Council (US). Washington: National Academy Press, 1982. Posted at http://www.nap.edu/catalog/10264.html) was based – in response to the assertion by DB Thomas (Echo correlation analysis and the acoustic evidence in the Kennedy assassination revisited. Science and Justice 2001; 41: 21–32) that alleged gunshot sounds (on Channel 1), apparently recorded from a motorcycle officer's stuck-open microphone, occur at the exact time of the assassination (as established by emergency communications on Channel 2). We have critically reviewed these two publications, and have performed additional analyses. In particular we have used recorded 60 Hz hum and correlation methods to obtain accurate speed calibrations for recordings made on both channels, cepstral analysis to seek instances of repeated segments during playback of Channel 2 (which could result from groove jumping), and spectrographic and correlation methods to analyze instances of putative crosstalk used to synchronize the two channels. This paper identifies serious errors in the Thomas paper and corrects errors in the NRC report. We reaffirm the earlier conclusion of the NRC report that the alleged "shot" sounds were recorded approximately one minute after the assassination.

Hemos revisado la evidencia acústica del asesinato de Kennedy-grabaciones de los dos canales de radio de la policía de Dallas, sobre los cuales se basó nuestro informe NCR original ((Ramsey NF et al., Report of the Committee on Ballistic Acoustics. National Research Council (US). Washington: National Academy Press, 1982. Posted at http://www.nap.edu/catalog/10264.html)-e en respuesta a la aseveración de DB Thomas ((Echo correlation analysis and the acoustic evidence in the Kennedy assassination revisited. Science and Justice 2001; 41: 21–32) de que los pretendidos sonidos de disparos (del Canal 1) aparentemente grabados por un micrófono de un oficial en motocicleta, ocurrieron exactamente al mismo tiempo que el asesinato (tal como se estableció en el Canal 2 de Comunicación de emergencias). Hemos revisado críticamente las dos publicaciones y hemos realizado análisis adicionales. En particular hemos usado hum registrados de 60 Hz y métodos de correlación para obtener calibraciones de velocidad para grabaciones hechas en los dos canales, análisis espectrales para buscar instancias de segmentos repetidos durante el play-back del canal 2(que pudiera resultar del salto de surco)y métodos de correlación y espectrográficos para analizar instancias de posibles conversaciones cruzadas usadas para sincronizar ambos canales. Esta publicación identifica graves errores en la publicación de Thomas y corrige errores del informe NCR. Reafirmamos la primitiva conclusión del informe NCR de que los pretendidos disparos de arma fueron registrados un minuto aproximadamente después del asesinato.

Als Antwort auf die Behauptung von D. B. Thomas (*Echo correlation analysis and the acoustic evidence in the Kennedy assassination revisited. Science and Justice 2001; 41:21-32*), dass angebliche Schussgeräusche (auf Kanal 1), die anscheinend über das offene Mikrophon eines auf einem Motorrad befindlichen Polizeibeamten aufgezeichnet wurden, genau zum Zeitpunkt des Attentats auftreten (wie durch Notrufgepräche auf Kanal 2 festgestellt wurde), haben wir im Mordfall Kennedy die akustischen Beweismittel der beiden Funkkanäle der Polizei von Dallas, auf der unser ursprünglicher NRC Bericht (*Ramsey NF et al., Report of the Committee on Ballistic Acoustics. National Research Council (US). Washington: National Academy Press, 1982. Posted at http://www.napo.edu/catalog/10264.html*) basierte, nochmals untersucht. Wir haben diese beiden Publikationen kritisch überprüft und haben zusätzliche Analysen vorgenommen. Im Einzelnen haben wir aufgenommenes 60 Hz Netzbrummen und Korrelationsmethoden benutzt, um genaue Geschwindigkeitskalibrierungen für auf beiden Kanälen gemachte Aufnahmen zu erhalten, Cepstralanalysen, um Fälle wiederholt auftretender Segmente während des Abspielens von Kanal 2 (was durch Rillensprünge verursacht worden sein könnte) zu suchen, und spektrographische und Korrelationsmethoden zur Analyse von Fällen vermeintlichen Übersprechens, das zur Synchronisierung der beiden Kanäle benutzt wurde. Dieses Papier zeigt schwere Fehler in der Veröffentlichung von Thomas auf und korrigiert Fehler im NRC

***Author for correspondence**

Key words Forensic science, ballistic acoustics, assassination, JF Kennedy, DB Thomas

[1]Members of the 1982 NRC Committee on Ballistic Acoustics.

Bericht. Wir bestätigen nochmals die frühere Schlussfolgerung im NRC Bericht, dass die angeblichen "Schussgeräusche" ungefähr eine Minute nach dem Attentat aufgezeichnet wurden.

Nous avons passé en revue les indices acoustiques découlant de l'assassinat du Président Kennedy – des enregistrements de deux canaux de police de Dallas sur la base desquels notre rapport original NRC (Ramsey et al., Report of the Committee on Balistic Acoustics. National Research Council (US), Washington: Natonal Academy Press, 1982. Posted at http://www.nap.edu/catalog/10264.html) était basé – en réponse à l'affirmation par DB Thomas (Echo correlation analysis and the acoustic evidence in the Kennedy assassination revisited. Science and Justice 2001; 41: 21–32) qui prétend que des sons de tirs (sur le canal 1) apparemment enregistrés par un microphone enclenché sur la moto d'un policier sont entendus au moment exact de l'assassinat (tel qu'il a été établi sur des communications d'urgence sur le canal 2). Nous avons passé en revue de manière critique ces deux publications et nous avons procédé à de nouvelles analyses. En particulier, nous avons utilisé un enregistrement de bourdonnements à 60 Hz et des méthodes de corrélation afin d'obtenir des calibrations de vitesse sur les enregistrements faits sur les deux canaux, une analyse spectrale pour chercher des cas de segments répétés durant la répétition du canal 2 (qui peut résulter d'un saut de sillon) et des méthodes spectrographique et de corrélation pour analyser les cas d'entretiens putatifs utilisés pour synchroniser les deux canaux. Cet article identifie des erreurs sérieuses dans l'article de Thomas et corrige des erreurs du rapport du NRC. Nous affirmons à nouveau les conclusions précédentes du rapport NRC qui concluait que les sons de soi-disant coups de feu ont été enregistrés à peu près une minute après l'assassinat.

Introduction

At the time of President Kennedy's assassination a police department microphone was stuck open for about five minutes, and the sounds it picked up were transmitted and recorded on the Dallas Police Department (DPD) Channel 1 Dictaphone plastic belt recorder. The Warren Commission in 1964 knew of this recording, but based none of their conclusions on it. Fifteen years later the US House of Representatives Select Committee on Assassinations commissioned studies by Barger, Robinson, Schmidt, and Wolf (BRSW) [1] of Bolt, Beranek and Newman, and by consultants Weiss and Aschkenasy (WA) [2]. Their reports concluded that this Dictabelt bore the acoustical imprint of shots, and that there was a 95% probability that there was an additional assassination shot from the Grassy Knoll (GK) area near Dealey Plaza. The FBI disagreed with the finding of a shot from the Grassy Knoll.

The Department of Justice then requested that the National Academy of Sciences and the National Research Council (NRC) study the reports and make recommendations. The NRC committee in its 1982 report [3] criticized some of the statistical calculations in the BRSW report and, more importantly, studied evidence of the crosstalk from the Dallas Police Department (DPD) Channel 2 transmission (recorded on a Gray Audograph plastic disk) onto Channel 1 (recorded on a Dictabelt). Channel 2 was used by the motorcade, police chief, and sheriff. The NRC Committee concluded that the sounds attributed to "shots" occurred well after Dallas Police Chief Curry had broadcast "Go to the hospital" (hereafter called GO), and hence long after the assassination.

The NRC's conclusion was reached on the basis of measurements of time intervals on the two channels, and of two instances of crosstalk between the two channels that could synchronize their timing. One crosstalk, "Hold everything secure" (hereafter denoted "HOLD") occurs at approximately the same time as the alleged shots on Channel 1, and one minute after GO occurs on Channel 2. The other crosstalk, "You want me to still hold this traffic on Stemmons" (hereafter denoted "YOU") occurs more than two minutes after the alleged first shot. The NRC Committee [3] concluded from the YOU crosstalk that the alleged shot sounds occurred 20–30 s or more after Chief Curry broadcast "GO to the hospital."

Since the NRC report is now out of print we have placed a copy on the Web [3]. It contains a wealth of detail and rationale, and should be read in conjunction with the present paper. A published summary of the Committee's results is also available [4].

In 2001, Thomas [5] published a paper in which he criticized the NRC crosstalk analysis, argued that it is invalid to use the HOLD utterance to synchronize the two channels, and claimed that "the gunshot sounds occur exactly synchronous with the time of the shooting" (i.e., the time of the assassination). The NRC Committee ceased to exist after its report was written in 1982; at least two members have died, so that it is impossible for the NRC Committee to write a response to the Thomas paper. However, the present authors, who include four former members of the NRC Committee, have studied the Thomas paper and reexamined the NRC Report, and here submit our analysis of the timing of the alleged "shots." We reach the conclusion that the sounds alleged to be gunshots were recorded approximately one minute after the assassination. As we were preparing this report for publication, Steve Barber (referenced in ([3], p. 4)) called our attention to an independent analysis by Michael O'Dell [6], which reaches conclusions similar to ours.

This paper is organized as follows: The audio source materials, and the methods used to create the later-generation digitized tracks used in the study, are described. To determine the relation between the speed of each track and that of the original recording, a combination of AC hum analysis and spectrographic cross-correlation methods is described and applied. Next, several issues and points of contention that are specific to the timing analysis for one or more of the tracks are raised and resolved. To resolve those issues, mathematical methods including cepstral analysis and special spectrographic techniques are applied, and lead to additional new findings that are described. Several acoustic events that are putatively present on both channels are then analyzed, using spectrographic cross-correlation and other methods, to determine which events are actually simultaneous on both channels, and thereby fix the ordering of crucial events in real time. We discuss differences between the present analysis and those of the NRC report [3] and Thomas [5], identify errors in the earlier two papers, and reach conclusions regarding the acoustic evidence for the timing of alleged shots fired by a "second gunman."

Materials

The NRC report, the Thomas report, and this paper all depend, for timing information, on crosstalks from DPD Channel 2 onto the Channel 1 record. There is general agreement among analysts that the crosstalk originated when a police department loudspeaker reproducing the sound of Channel 2 was within pickup range of the stuck-open microphone or of a competing microphone broadcasting on Channel 1. The original Dictabelt recording of DPD Channel 1, which contains the alleged sounds of the "shots," and the original Gray Audograph disk, which recorded the broadcasts on DPD Channel 2 from the motorcade, are held by the National Archives and Records Administration (NARA) and are not generally available for study. Over time, copies with various degrees of availability and quality have been made.

In 1963, James Bowles, Communications Supervisor at the time of the assassination, made a copy of the Channel 1 recording using a Dictaphone for playback, and a copy of the Channel 2 recording using a Gray Audograph for playback. Bowles used acoustic transfer – the microphone of the tape recorder juxtaposed to the loudspeaker of the playback machine. The stylus on the Dictaphone is free-running on playback and no artifacts in the form of "repeats" have been detected on the relevant portions of the Bowles copy of Channel 1. In contrast, for the Audograph playback of Channel 2 the position of the disk relative to the fixed stylus is dominantly driven mechanically ([3], p. 61) and the many repeats and skips on the Bowles copy of Channel 2 made timing measurements uncertain.

In the presence of several members of the NRC Committee in 1981, Bruce Koenig of the FBI Technical Services Division copied the original Dictabelt and Audograph records onto 7.5 ips reel-to-reel tapes. (These NRC Committee copies produced at the FBI will be referred to for brevity as the "FBI copies" of Channels 1 and 2.) He also found that he could not play the original Channel 2 disk without repeats using a Gray Audograph playback. He then used a free-running phonograph turntable for playback and provided the NRC a magnetic-tape copy of Channel 2, recorded at 7.5 ips, having no apparent repeats within the relevant sections. The turntable playback at constant angular velocity of the Audograph inside-out disk results in a pitch that steadily increases throughout the tape copy. Depending on the nature of the rectifier circuit used in the Dictaphone or Audograph, a dominant AC hum might initially have been recorded at 60 or 120 Hz.

In this report we use eight audio tracks that were generated from the Bowles and FBI recordings. We have digitized our best copies to CD-format WAV files (16-bit, at 44,100 samples/s), reproduced the tracks on two audio CDs, and (for ease of download by others) compressed the seven tracks of CD #1 and Track 6 of CD #2 (denoted here as "Track 6B") to corresponding MP3 files, which are freely available via Web server [7]. For the tracks of CD #1: Track 1 is the relevant section of the Bowles Channel 1 recording; Tracks 2 and 3 are two successive segments of the Bowles Channel 2 recording; Track 4 holds miscellaneous short segments; Track 5 is from the FBI playback of Channel 1; Track 6 was made from the FBI turntable playback (33 1/3 rpm) of Channel 2, recorded onto tape at 7.5 ips by the FBI, and played back at 7.5 ips; and Track 7 is that same recording, played back at 3.75 ips. Track 6B includes a portion of the Bowles Channel 1 recording that occurs following the end of Track 1. It was transferred to cassette by one of the authors (NFR) in 1983 from a Bowles copy, and was digitized and transferred to CD in 2001.

Owing to the method by which Tracks 6 and 7 were created, the pitch of utterances increases with time on both tracks. Over the time interval of interest the frequencies on Track 7 are closer to those of normal speech; therefore Track 7 is used in our analysis, and will be referred to as the FBI copy of Channel 2. For Track 7, we established by cepstral analysis (described below) that the transcription from magnetic tape copy to digitized CD audio contributed negligible additional timing error (combined wow, flutter and speed error less than 0.2%). We assume that the transfer from our tapes to the digital files is of the same quality for the other tracks.

Methods

Reconstruction of timelines

Reconstruction of the timelines involves several steps. As noted, each sound track used in our analysis was created by playing back and copying one of the original source media (Dictabelt or Gray Audograph), in some cases through several generations. We distinguish three types of event timings. "Playback time" denotes the measured time at which an event occurs on one of these tracks (before making any corrections for recording speed or for possible skips, repeats, and recorder stops). Note that the speed at which the source medium was originally recorded may be different from that at which it (and/or an intermediate recording) was played back to generate a track used in our analysis. By "recording time" we mean the cumulative time for which the original recorder had been running at the time of the event. We compute an estimate of the recording time of an event from the measured "playback time" by correcting for the differences between recording and playback speeds. To do this, each increment of playback time is multiplied by a speed correction factor K to obtain the corresponding increment of recording time. We use the AC hum that is present on these recordings, in conjunction with the method of spectrographic "pattern cross-correlation" discussed below, to determine the values of K. Note that the computed recording time will differ from the true recording time if skips or repeats occurred during any of the playbacks and re-recordings that generated the track in question. In this paper all times marked with subscript 'r' denote the computed recording times using these K values.

It is known that the Gray Audograph was prone to skip and repeat grooves on playback. Owing to the construction of the Audograph, forward jumps (skips) and backward jumps (repeats) of the stylus on playback must approximately cancel each other out over time. That is, the net sum of the number of skips minus repeats, at any point in the recording, must remain small. This

Table 1 Playback and recording times for several phrases on Channels 1 and 2

A PHRASE	B t_{7pm}	C t_{7rs}	D t_{2pm} $\{t_{3pm}\}$	E t_{2rs} $\{114.7 + t_{3rs}\}$	F t_{1pm}	G t_{1rs}	H t_{5pm}	I t_{5rs}
CHECK	12:39	−99.0	2:07	−95.2				
12:30	12:51.5	−87.5	2:17.6	−83.9				
GO	13:13.1	−67.6	2:36.3	−63.9				
Loud Go	13:18.7	−62.4	2:41	−58.9				
CHECK1					3:45	−12.4	3:05	183.2
Bell-a	14:13.8	−10.7						
"First shot"					3:50.9	−6.5	3:09.6	187.7
HOLD	14:25.1	0.1	3:36.4	0.4	3:57.4	0.1	3:16.1	194.1
Bell-b	14:33.2	7.8			4:05.1	7.9	3:24.1	202.1
12:32	14:59.9	33.6						
Dispatcher	15:45.5	78.3	4:54	83.5				
Tape break			5:19.1{0:00}	110.3				
First 15 car 2	16:20.8	113.4	{0:03.9}	{118.6}				
12:34	16:31.7	124.4	{0:18.3}	{133.2}				
YOU	16:50.3	143.2	{0:33.4}	{148.4}	6:48.6	173.0	6:12.4	368.7
12:35	17:27.9	181.7	{1:11.2}	{186.6}				
12:35 (repeat)			{1:14.9}	{190.3}				
12:35	17:58.2	213.2	{END}					
12:36	18:26.5	242.9						
12:36	18:56.6	274.9						

(A) Acoustic event. (B) Playback time (mm:ss) on Track 7 (all tracks refer to audio CD #1). (C) Track 7 recording time (s). (D) Track 2 playback times (mm:ss) down to "tape break" and Track 3 (bracketed entries) for rest of column. (E) Track 2 recording time (s) down to tape break, and Track 3 (bracketed entries) for rest of column. This procedure is intended to make the entire column equivalent to what the recording times would have been if there had been no tape break (see text). (F) Track 1 playback times (mm:ss) for Bowles copy of Channel 1. (G) Track 1 recording times (s). (H) Track 5 playback times (mm:ss) for FBI copy of Channel 1. (I) Track 5 recording times (s).

problem gives rise to an added uncertainty that is probably no greater than 8 s for the measurement of any interval of time on Tracks 2 and 3. The time intervals calculated from Tracks 2 and 3 in Table 1 are computed on the assumption that there are no unbalanced skips.

"Actual time" denotes the time at which an event actually occurred. Even after the speed correction is correctly made, and after possible skips or repeats are accounted for, an interval of recording time is not necessarily equal to the actual time elapsed, since there may have been some dead time during which the recorder was stopped. Both channels of the DPD recording systems were signal activated; i.e., if the transmission stopped for more than 3 or 4 s, the recorder stopped recording until a new transmission occurred. Since a transmitter on Channel 1 (with the stuck-open microphone) was continuously transmitting during the relevant period, there was no dead time (even though there were long stretches with no voice), so the recording time on Channel 1 was the actual time elapsed. But Channel 2 may have had some dead time, and thus increments of the recording time on Channel 2 may be less than the actual time elapsed. This dead time necessarily exists on all the copies of Channel 2, preventing one from directly measuring the actual times. In addition, on the Bowles copy of Channel 2 (our Tracks 2 and 3), a tape break occurred during the time interval of interest. To infer both dead time and the (short) duration of this tape break, we use evidence based on synchronization of several acoustic events between Channels 1 and 2.

In this report we denote playback time by t_p and recording time by t_r. An additional numerical subscript denotes the track number, and a final subscript m denotes that the times are marked in minutes and seconds. For example, t_{7pm} indicates the playback time in minutes and seconds as heard on Track 7. The subscript m will be replaced by s to indicate that the time is in seconds and that the time origin has been shifted by subtracting a fixed time, which is 3:57.3 or 237.3 s for Track 1, 3:36 or 216 s for Track 2, 14:25 or 865 s for Track 7, and 0 s for the other tracks. Whenever we refer to an acoustic event on a track by giving only its time, it is the playback time t_p on that track that is being referenced; i.e., the location of the event in the files of [7].

Speed correction factors K

The logic by which we infer the speed correction factors (denoted K_n for Track n) is as follows. The AC hum on Track 7 (Channel 2) is used to determine the time-dependent factor $K_7(t_{7p})$. The "pattern cross-correlation (PCC)" method (below) is then used to compare the same utterances on short sections of two different tracks, and to determine the speed ratio between those sections for each utterance (we call these pairs "time ties"); this yields the ratio of K values for the two tracks at the time ties. The two tracks may be recordings of the same channel, or of different channels in cases for which the utterance is clearly a broadcast (on both channels at once) or a crosstalk (from one channel to the other). In order to determine K for the entire track, rather than only at the time ties, we examine hums that persist throughout the entire recording on each track. For all tracks except Track 7 (and 6, which we do not use), the hum frequencies are quite constant, showing that K for those tracks is essentially constant in time. We thereby obtain a consistent set of K values for all tracks of interest (Tracks 1 and 5 for Channel 1, and Tracks 2, 3, and 7 for Channel 2).

Uses of AC hum for speed determination

We cannot directly use the hum frequencies to determine K for Tracks 1, 2, or 3 (without also using PCC), because there are multiple hums that are not all harmonics of a single frequency (as shown in "Results" below). These multiple hums may have been introduced during multiple generations of re-recording, and/or by electromechanical or acoustic noise that, while driven by 60 Hz AC, does not itself have a frequency of 60 Hz. [For example, acoustic noise from a fan operating at, or slightly slower than, a standard rate of 1725 rpm can generate a hum at or slightly lower than 57.5 Hz, and it is known that the Bowles recordings (Tracks 1, 2, and 3) were made by acoustic transfer in open air, rather than by an electrical coupling.] Therefore, one cannot reliably use the hums to determine K values by assuming that a particular hum was present on the original recording (rather than introduced later) and had a frequency of 60 Hz when originally recorded. However, one can use these hums (independent of their source, provided the source frequency was constant) to show that the K values for Tracks 1, 2, and 3 are constant over time. In the case of Track 5, the FBI states that the speed of this recording was regulated by the observed AC hum (in this case, at 120 Hz). As we shall see, there is indeed only one prominent hum on that track; it is at 120 Hz; and the K value of 1.00 implied by that finding is consistent with the K values obtained by the PCC method comparing Tracks 5 and 7 at time ties.

Track 7 (as well as 6) is distinctive in that it has an AC hum that can definitively be identified as the hum that was introduced when the Gray Audograph (used for Channel 2) was played back to create the "FBI copies." This is true because the Gray Audograph (used for Channel 2) was designed to record at nearly constant *linear* speed along the spiral track (from the inside out), as contrasted with the conventional vinyl recording technique of constant *angular* velocity (i.e., constant rpm). However, to avoid skips or repeats, the FBI used for playback a standard phonograph turntable revolving at constant angular velocity. As a consequence, when the tape is played back, the ratio K of recording time interval to playback time interval increases linearly with playback time for Tracks 6 and 7. Therefore ([3], p. 68),

$$K(t_{\mathrm{ps}}) = a + b\,t_{\mathrm{ps}}.$$

The recording time is therefore

$$t_{\mathrm{rs}} = \int K(t_{\mathrm{ps}})\,dt_{\mathrm{ps}} = a\,t_{\mathrm{ps}} + bt_{\mathrm{ps}}^2/2$$

where a and b are constants to be determined (and where t_{rs} is defined to be zero when $t_{\mathrm{ps}} = 0$). Owing to the offset of the Audograph's rubber drive wheel from the radial position of the recording or reading stylus, the linear speed along the track is not constant. Nevertheless, an analysis of the Audograph mechanism shows that the expressions for $K(t_{\mathrm{ps}})$ and t_{rs} remain precisely of the form given above.

The unique hum on Track 7 that varies linearly with playback time must be one that was present on the original Gray Audograph medium; any hum added during the FBI copying process (or later) would not show the linear variation. Therefore $K_7(t_{7p})$ can be determined absolutely, on the assumption that the original hum was indeed at 60 Hz. In fact, even if the original hum were at some other frequency (for which no evidence has been presented), this would only imply that the K values for *all* tracks need to be multiplied by the same factor. It would in no way affect the *relative* speeds of the various tracks, and would therefore leave unchanged all conclusions concerning the sequence of acoustic events on the two channels.

Pattern cross-correlation method

This method is used to confirm the presence and determine the precise relative timing of Channel 2 broadcasts that have been picked up as crosstalk and recorded onto Channel 1 along with other sounds in the vicinity of the Channel 1 microphone. The Channel 2 waveform (as recorded onto Channel 1) is subjected to frequency- and time-dependent modulation, distortion, and (as a result of Channel 1 automatic gain control) nonlinear attenuation. Therefore the usual method of signal cross-correlation, in which one seeks a peak in the cross-correlation between waveforms obtained from the two channel recordings, will not provide a reliable signature for the presence or relative timing of the crosstalk. Instead, we compute spectrograms of (a) the desired short segment of Channel 1 and (b) of a longer Channel 2 segment, a portion of which was putatively responsible for the crosstalk heard on Channel 1. We then use a "pattern cross-correlation (PCC)" method, described below, to measure the presence of correlations between energy-containing regions in the two spectrograms, and to determine the timing offset between the correlated regions. In addition, we use PCC to determine the relative speeds of the two recordings, by finding the relative speed correction (which affects both frequencies and time intervals) for which the PCC exhibits the strongest peak. (This general approach was described in [3] and more fully in [8].)

To compute the spectrograms, each waveform is digitally lowpass-filtered to 3500 Hz, then resampled at 8820 samples/s. Each frame is 512 samples long and is shifted by 64 samples relative to the previous frame, i.e., overlapped by 448 samples. The waveform within each frame is multiplied by a Hamming window, and the fast Fourier transform (FFT) is computed. The square of the absolute value of the FFT yields a value of power at each of 512 frequencies (each frequency "bin" is 17.23 Hz), and at a number of times that equals 1/64 the number of sample points in the waveform. In the plots (except where otherwise stated), the intensity denotes the spectral power density (at each time and frequency) raised to the 0.3 power, so that the large range of spectral power density should be visible to the eye.

To compute the PCC, the power in the two spectrograms is cross-correlated at each frequency. Each such cross-correlogram is a function of the relative time shift between the channels. The value of the PCC at each time shift is obtained by summing the cross-correlogram values (at the same time shift) over frequency, applying a weighting factor that boosts the contribution of higher frequencies to the sum by 3 dB per KHz, and in some cases normalizing for power. A sufficiently strong and clear peak in the PCC provides evidence that the Channel 1 segment is present within the Channel 2 recording at a relative time shift given by the position of the peak along the time axis. The PCC peak is diminished if one of the channels is sampled at the wrong rate (speed). By repeating the PCC calculation, varying the speed of Channel 1 before computing its spectrogram, and comparing the height of the PCC peak for each speed correction, one can obtain an indication of the relative speed of the two channel recordings. The position of the peak in time indicates the relative timing (i.e., the time offset as opposed to the relative speed) of the utterances to about 0.01 s. Thus the PCC determines both the relative timing and speed of the two channels, and can confirm the presence of putative crosstalk.

The behavior of the PCC peak, when time and frequency "warps" (compression or expansion factors) are introduced, can reveal information as to whether the peak is indeed a signature of the identical utterance recorded on two channels. An increase in speed by a given factor, of course, decreases time intervals and increases frequencies by that factor. If the same utterance has indeed been recorded on two channels at different speeds, the resulting PCC would typically be expected to have a peak when the appropriate speed correction factor is applied. Note, however, that one can also mathematically apply independent warps to the time and frequency axes separately. If a PCC peak is an indicator of the same utterance having been recorded at different speeds, then the time and frequency warp factors at which that PCC peak achieves its maximum should be the inverse of one another. On the other hand, if the PCC peak is maximized when the time and frequency warp factors are not inverses of one another, this will suggest that the PCC peak is not a signature of a single utterance having been recorded at two different speeds. This method of analysis will be applied (below) to three putative crosstalks that have been used in different ways by various workers to synchronize the timings between Channels 1 and 2. To summarize, we compute the speed warp (which compresses or expands time and frequency in a reciprocal way) that maximizes the PCC peak, and then also compute the variation of the PCC peak as a function of the "duration-only" warp (which compresses or expands the time axis if the warp factor differs from unity, but does not affect frequencies. A "duration-only" warp different from a value of unity would not physically occur; therefore, if the putative crosstalk is genuine, the PCC peak should be maximized when the "duration-only" warp is close to unity.

Cepstral analysis and detection of skips and repeats

The cepstrum [9], defined here as the inverse Fourier transform of the logarithm of the magnitude of the Fourier transform of the signal, was originally devised to facilitate the detection of echoes in acoustic signals. The input to this function is a time-varying signal that may contain a component that is repeated or added, possibly with attenuation, with some time delay. The output value of this mathematical function has a sharp maximum at a time that corresponds to the time interval between the two occurrences of the repeated signal component, if such a repeat is present. Cepstral analysis is well suited to detect possible repeats that might result from a phonograph stylus jumping to a previous portion of the groove during the FBI playback of the Gray Audograph (Channel 2) disk, which was used to create Tracks 6 and 7. On Track 7 a repeat that results from a groove skip-back of one turntable revolution should correspond to a strong cepstral peak at a signal delay that corresponds to the revolution time of approximately 1.8 s (33 1/3 rpm). For convenience of analysis, we digitally lowpass-filtered the Track 7 data, keeping frequencies up to 2000 Hz, then downsampled the filtered signal from 44100 to 4410 samples/s. Since Track 7 was recorded from the FBI playback at half speed, the apparent repeat time on Track 7 should be approximately 3.6 s, corresponding (at 4410 samples per second) to a sharp cepstral peak occurring at a delay of approximately 15,876 samples.

"Gabor Spectrogram" Analysis

To obtain an additional useful pictorial representation of selected utterances and other acoustic signals, we use a method [10] in which a special type of spectrogram, which we refer to here as a "Gabor spectrogram," is generated. The given waveform is processed by a bank of overlapping narrow-bandpass digital filters. The center frequencies of each bandpass filter are equally spaced in the logarithm of frequency, and the width of each filter (specifically, the full width at half maximum) is a constant fraction (approximately 3%) of the center frequency. In the frequency domain, each filter is a "log-Gabor" filter; that is, a Gaussian function of $\log(f/f_{\mathrm{center}})$. [For narrow bandwidth this is similar to a Gaussian function of $(f - f_{\mathrm{center}})$, which is the usual definition of a Gabor filter.] The output of the filterbank is a set of complex numbers, one for each band at each discretized value of time. The "Gabor spectrogram" plots the absolute magnitude of these output values as a function of $\log(f)$ and time, where the frequency ranges from 160 to 5120 Hz. The particular choice of the filter width in the frequency domain (as a fraction of center frequency) yields, for speech signals, a Gabor spectrogram in which the output is "sparse," that is, the fraction of the $\log(f)$ vs. t plane having significant signal power (or output value) is small. If the filter width were much increased, the signal power would be spread over a larger range of frequencies, so the frequency resolution would be degraded; if the width were much decreased, the time resolution would be degraded. Unlike a conventional spectrogram, in which the size of the time frame (e.g., 10 ms) for each computation of the power spectrum is the same for all frequencies, the Gabor spectrogram effectively applies a shorter time window at high frequencies than at lower ones, enabling the resolution of the signal power in both the frequency and time domains to be jointly optimized.

Results

Issues that affect the interpretation and synchronization of the various recordings include possible repeats, possible skips, determinations of the speed correction factors K, instances of crosstalk, interpretations of spoken words, and possible incompatibility with the dispatcher's time annotations. (The "overdub" hypothesis asserted by Thomas [5] to invalidate the use of the HOLD utterance for synchronization will be analyzed later, in the "Discussion" section.)

This section is organized as follows. The speed correction factors K, which relate recording to playback time for each track, are derived from analysis of Track 7 AC hum and PCC "best warp" analysis of simultaneous utterances on different tracks (corresponding to either the same or different radio channels). Our summary of the timing of key acoustic events on each channel (Table 1) is described. Specific timing issues are then discussed for each track in turn. For Track 7 (obtained by playing the Audograph recording on a standard turntable) we present results on whether (and, if so, where) that track contains repeats (skips backward) or skips forward, and we reconcile apparent inconsistencies between Track 7 and Track 2 (which was created by playing the Audograph recording back on an Audograph machine, in which case skips and repeats are known to occur). We present a new finding, that of "premonitory whisper repeats," which are found by cepstral analysis and confirmed by direct listening. For Track 5 (the "FBI copy" of Channel 1) we find a prominent speed irregularity or "warble" by means of spectrographic analysis. We then consider three putative crosstalks ("You want me . . .," "Hold everything secure," and "I'll check it"), and determine which of these are valid crosstalks that can be used to synchronize the two channels, by means of cross-correlation and other methods. Finally, we determine the timing relationship between the utterance "Go to the hospital" (which immediately follows the actual assassination shots) and the acoustic events alleged to be the imprint of gunshots. Our rebuttal of Thomas' argument concerning the dispatcher's spoken time annotations appears in the "Discussion" section below.

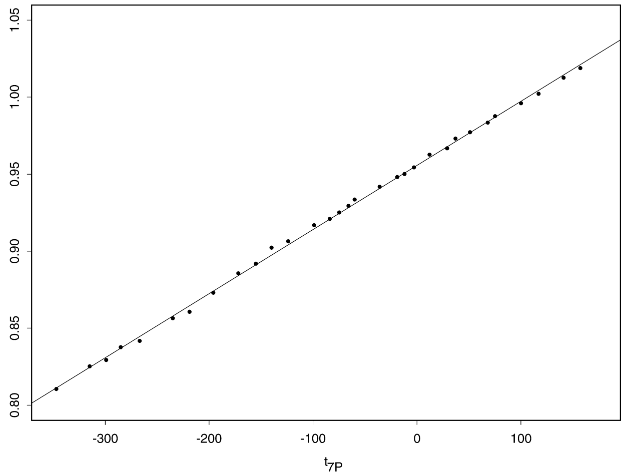
Speed correction factors K

For Track 7, our evaluation of K is based on spectral analyses of the recorded 60 Hz hums on the digital recordings from 08:38 ($K = 0.811$, nearly 6 min before the utterance "Hold everything secure," or HOLD) to 17:02 ($K = 1.021$, 12 s after the utterance "You want me to still hold this traffic on Stemmons," or YOU). The CD audio was input to an SRS785 digital spectrum analyzer, with 8-s integrations and Blackman–Harris window.

For many time windows within which a clean AC hum feature was found on Track 7, we measured the frequency of the AC hum, and computed the resulting value of the speed correction factor K. These values of K are plotted in Figure 1 with the corresponding linear least squares best fit, which is: $a = 0.9556 \pm 0.0004$; $b = 0.000416 \pm 0.000006$. As noted, for Track 7 there can be no doubt that the hum was recorded along with the original sound, and not during any subsequent copying process.

As noted in the section "Methods: Uses of AC hum . . ." above, and as shown in Figure 2, the multiplicity of "hum" lines in the spectrograms (not corresponding to a single fundamental frequency) for Tracks 1, 2, and 3 preclude our unambiguously identifying one of them as corresponding to an original 60 Hz AC hum or one of its harmonics. We therefore use PCC to determine K ratios, in two ways. First, where the same acoustic event occurs on two tracks, we adjust the relative speed by a constant factor in the vicinity of that event (i.e., "warp" the speed of one of the tracks) until the PCC peak is maximized (we call this the "best warp"). Second, we survey a long time interval (e.g., 3 min) and note how the position in time of the PCC peak (i.e., the time shift between the corresponding acoustic events) "creeps" as time advances; this yields both an average ratio of K values over the time interval, and a measure of the constancy of that ratio. Finally, we note that the constancy in time of the hums in Figure 2 for Tracks 1, 2, 3, and 5 shows that the locally obtained "best warp" K values for those tracks are indeed essentially constant over the several-minute region of interest (assuming that the sources of those hums were at constant frequencies).

Figure 1 Time correction factor K versus playback time $t_{7\text{ps}}$ for Channel 2 Track 7. Time is in seconds with origin at CD playback time 14:25.

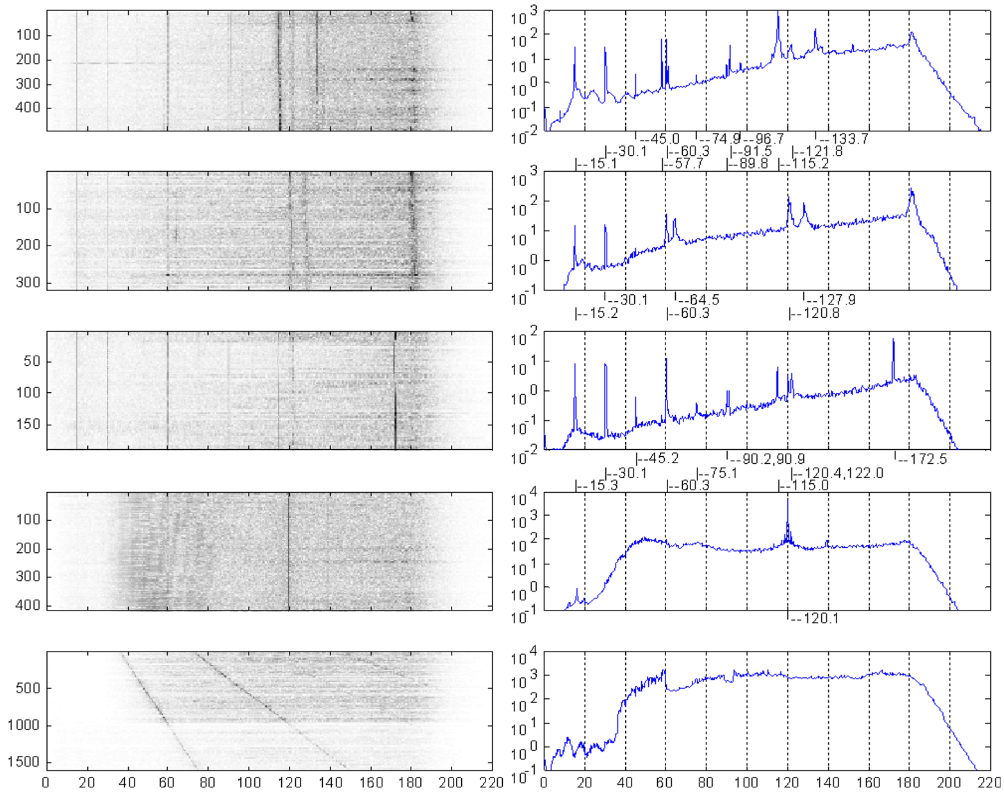


We first consider "best warp" PCC of Track 2 vs. Track 7 (both corresponding to Channel 2) at the utterance HOLD. This PCC (not shown) has a clear peak whose magnitude is greatest when the speed warp corresponds to a ratio $K_2/K_7 = 1.120 \pm 0.001$; and since $K_7(\text{HOLD}) = 0.9556$, therefore $K_2(\text{HOLD}) = 1.07$.

An utterance referred to as [PL]AY, and discussed below in more detail, occurs at $t_{2\text{pm}} = 4{:}59.5$ and on Track 7 in the vicinity of $t_{7\text{pm}} = 15{:}56$. "Best warp" PCC yields $K_2/K_7 = 1.08$; since $K_7([\text{PL}]\text{AY}) = 0.9556 + 0.000416 \times (956 - 865) = 0.9935$, we have $K_2([\text{PL}]\text{AY}) = 1.07$, consistent with the value of $K_2(\text{HOLD})$ above.

Next we compare Track 1 (Channel 1) and Track 7 (Channel 2) at the putative crosstalks HOLD and YOU. (We later discuss the evidence that these are indeed valid crosstalks.) "Best warp" PCC yields $K_1/K_7(\text{HOLD}) = 1.055$; combining this with $K_7(\text{HOLD}) = 0.9556$ yields $K_1(\text{HOLD}) = 1.008$. The same comparison at the putative crosstalk YOU yields: $K_1/K_7(\text{YOU}) = 1.000$; and we know $K_7(\text{YOU}) = 1.0160$; therefore $K_1(\text{YOU}) = 1.016$. Both K_1 values are consistent with 1.01 to within experimental error. Note also that, by direct

Figure 2 Hum spectrograms and power plots. Left column: The short-term hum spectrum is displayed graphically for Tracks 1, 2, 3, 5, and 7 (top to bottom). The WAV files at 44100 samples/s were lowpass filtered to 800 Hz using GoldWave, and then resampled to 2205 samples/s. Each was then decimated in Matlab to 441 samples/s, using a preceding eighth-order Chebyshev type-I lowpass filter with cutoff frequency 176.4 Hz. The x axis is the frequency in Hz, and the y axis is playback time in seconds. To compute the spectrogram, a frame of 2048 samples is stepped 512 samples at a time, each frame is multiplied by a Hamming window, and the square of the absolute value of the relevant FFT component (raised to the 0.3 power to increase the range of values that is visible in the image) is plotted. The image for Track 1 shows multiple hum lines, presumably from the initial Dictaphone recording, from the Bowles acoustic transfer to tape, and perhaps from later tape-to-tape copying. The Track 2 and 3 images show hum lines from the original Audograph recording, from the Bowles acoustic transfer to tape, and perhaps from later tape-to-tape copying. The Track 5 image shows hum lines from the FBI transfer of the Dictabelt, showing a precise 120-Hz hum and no obvious additional hum signals. The Track 7 image shows hum lines for the Audograph disk recorded by FBI on tape from an audio turntable. The process evidently added no perceptible hum (which would be at constant frequency if present). The original Audograph hum components increase in frequency linearly with playback time. Right column: A plot of the short term summed power (y axis, log scale) in each of the 1024 frequency components from 0 to 220.5 Hz (x axis). For each component, the y axis is the sum (over frames) of the squared absolute FFT component values that were used to generate the spectrograms in the left column. Numeric labels beneath each subplot indicate the frequencies (in Hz) of identified hum lines.



listening, the playback time interval from HOLD to "Bell-b" (see Table 1) is 8.1 s on Track 7 and 7.7 s on Track 1. Since K is inversely proportional to the playback time interval (for a given recording time interval), we have $K_1/K_7 = 8.1/7.7 = 1.05$, hence $K_1 = 0.9556 \times (8.1/7.7) = 1.01$.

An additional speed comparison between the two channels is provided by the simultaneous broadcast (on both channels) "Attention all emergency vehicles . . . Do not use Industrial Boulevard," which appears in the Bowles transcript at 12:36 P.M. This utterance occurs on Track 7 (Channel 2) during the interval 18:18 to 18:29. The Channel 1 recording of this utterance is unfortunately not present on Track 1 (which ends before the utterance), but it is present on a different track (Track 6B) during the interval 12:43 to 13:02. Also, the earlier YOU utterance on Channel 1 is present on both Tracks 1 and 6B, allowing the relative speed of those two tracks to be determined. The results are as follows.

PCC "best-warp" comparison of YOU between the two Channel 1 copies yields $K_{6B}/K_1 = 0.980$. PCC comparison of ATTENTION between Track 7 (Channel 2) and Track 6B (Channel 1) yields $K_{6B}/K_7 = 0.943$. By the Track 7 calibration, K_7(ATTENTION) $= 1.0534$. Therefore K_{6B}(ATTENTION) $=$ 0.991, yielding $K_1 = 0.991/0.980 = 1.01$. Thus PCC analyses of three separate utterances yield the same value of $K_1 =$ 1.01 ± 0.01.

Prior to the above analyses for Channel 1, and when there was only a single measurement of K_1/K_7(HOLD), Thomas [5] correctly pointed out that the value of K_1/K_7 measured during a three-second interval at HOLD might be quite different from the average value of the speed over a much longer time if the recorder fluctuated badly. However, with K_1 derived from measurements at HOLD, YOU, and ATTENTION all agreeing within 1% it is highly unlikely that the value of K_1 would vary widely within this interval, and there is indeed no evidence favoring such a variation. Furthermore, as noted above, the constancy of hum frequency on Track 1 provides evidence for the constancy of K_1 throughout the time interval of interest.

Comparing Tracks 1 and 5 (both of Channel 1): At the "CHECK1" utterance (see Table 1 and discussion below), "best warp" PCC gives $K_1/K_5 = 1.038$. For the 200 s-long interval running from a few seconds before CHECK1 to after "YOU ... Stemmons," the "time creep" of the PCC peak yields $K_1/K_5 = 1.028 \pm 0.001$. Combining this with $K_1 = 1.01 \pm 0.01$ yields $K_5 = 0.98 \pm 0.01$. Note that there is only one prominent hum spectral line for Track 5 (Figure 2), at 120.0 Hz, and that this is consistent with the FBI's statement that Track 5 was recorded by setting the playback speed so as to place the AC hum at (a harmonic of) 60 Hz. This implies $K_1 = 1.00$, close to the results of our PCC measurements. For our calculations we use the intermediate value $K_5 = 0.99 \pm 0.01$.

Finally, comparing Tracks 3 vs. 7 (both Channel 2) at YOU using "best warp" PCC yields K_3/K_7(YOU) $= 1.000$; since K_7(YOU) $= 1.016$, we find K_3(YOU) $= 1.016$. Track 3 has a short overall duration, so the timing of key utterances is relatively insensitive to the precise value of K_3.

Based on the above evidence, we use throughout this report the values $K_1 = 1.01$, $K_2 = 1.07$, $K_3 = 1.01$, $K_5 = 0.99$, each constant in time and with an ascribed error of ± 0.01. K_7 is given by the linear relation $K(t_{ps}) = a + bt_{ps}$, where $a = 0.9556 \pm 0.0004$ and $b = 0.000416 \pm 0.000006$.

Figure 2 shows that, for each of Tracks 1, 2, and 3 (and for Track 5 as already noted), there is a spectral line corresponding to an original AC hum frequency of 60 Hz or one of its harmonics, when the above K values computed using PCC and the known function $K_7(t)$ are used. There are, as noted, other hum lines at frequencies that are not harmonics of an original 60 Hz AC hum; in several cases these correspond, after K speed correction, approximately to 57.5 Hz (or to a slightly lower frequency) or to one of its harmonics, which (as we have speculated above) may be the result of acoustic fan noise or machinery rumble introduced during the recording of Tracks 1, 2, and/or 3.

Summary of derived event timings
The key issue to be resolved is the relation between the actual time of the GO utterance and the occurrence of the alleged first shot. GO was heard only on Channel 2, and the "shot" only on Channel 1. We will synchronize the timelines for the two channels by using either or both of two instances of crosstalk: the utterances HOLD and YOU. The two synchronizations must be consistent with one another, although they might differ, and one of them might place much tighter bounds on timing than the other.

First, we determine the recording times of key acoustic events on each channel as summarized in Table 1. Column A lists various key phrases (among them, dispatcher's time annotations) and acoustic events. Column B provides the playback times t_{7pm} as given by the recording for Track 7, while Column C gives t_{7rs}, which is the computed recording time in seconds. Columns D and E are analogous to B and C, with the first portion of each column being for Track 2 (Bowles Channel 2 before tape break), and the second portion (in curly brackets) being for Track 3 (Bowles Channel 2 after tape break). Time is added for the tape break as discussed below. The adjustments are such that all of Column E should give the recording times as they would have been had there been no tape break. The listed values in Columns D and E are based on the assumption that the accumulated skips and repeats were balanced at the time the playback time was read. Columns F and G give the playback and computed recording times t_{1pm} and t_{1rs}, respectively, for Track 1, the Bowles copy of Channel 1; and Columns H and I present playback and computed recording times for Track 5, the FBI copy of Channel 1. The playback times in Table 1 can easily be checked using the recordings on the CD or on the Web. The recording times can readily be obtained from the playback times using the time correction formulas and offset constants given above.

Analysis of tracks 7 and 2 timing (Channel 2)
for possible skips and repeats
To compute Channel 2 recording times from the measured playback times on Track 7, we use (a) the speed correction K as derived above, and (b) the evidence, based on cepstral analysis as well as direct listening, that Track 7 contains no repeats (groove skipbacks) within the interval of interest (from GO at 13:13.1 to YOU at 16:50.3).

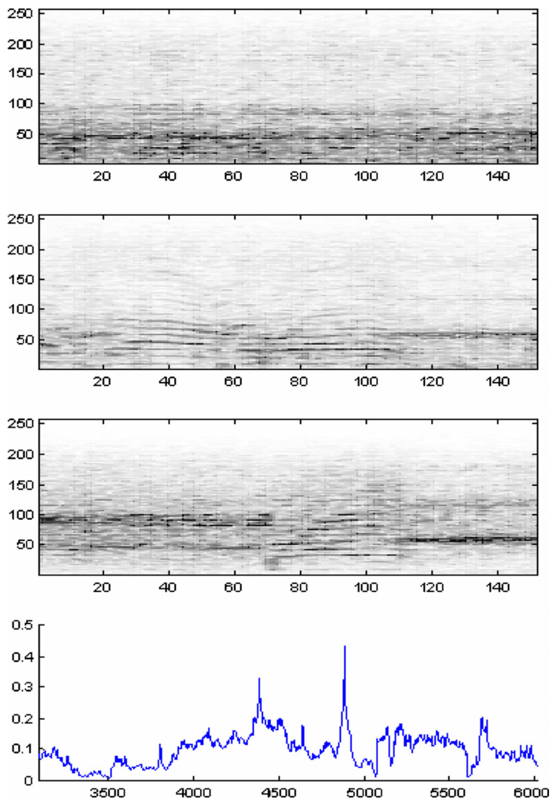
We searched for both forward skips and repeats in the Track 7 data. Thomas (personal communication, 2002) has claimed that there are at least two utterances that are present on Track 2 but absent from the NRC/FBI phonograph playback (a portion of which constitutes our Track 7), indicating the existence of forward skips on the phonograph playback. We analyze both of these cases (denoted below as "[PL]AY" and "Stand by"), as well as another utterance ("15 car 2 ... now... on Main"), that raise questions regarding the integrity of Track 7 timing. We find in each case that there is no Track 7 forward skip. In addition, the cepstral analysis rules out any Track 7 repeats indicative of a backward skip. Cepstral analysis also reveals a phenomenon that we call "premonitory whisper repeats," which do not reflect a timing problem (in fact, they provide a method for confirming the regularity of Track 7 timing to high accuracy), but are a by-product of the physical process of recording on a Gray Audograph disk. In contrast to Track 7, Track 2 is known to contain numerous forward and backward skips, which approximately compensate each other over time, owing to the construction of the Gray Audograph machine that was used for the playback that created Track 2.

The utterance "[PL]AY": Regarding a supposed Track 7 skip, Thomas (personal communication, 2002) has claimed that there is an utterance "twenty" on Track 2 (during the interval between GO and YOU) that is not apparent on Track 7. (The utterances discussed in this section appear in the Bowles transcripts [11].) There is in fact, on Track 2, a very short sound at 4:59.5 (sounding to us like the word "play," rather than "twenty") that is not recorded on Track 7 (it would be expected to occur at about 15:57). The utterance on Track 2 occurs against a simultaneous background of noise sounding like screeching tires and/or sirens. We have performed cross-correlation analysis between the corresponding section of Track 7 (which also contains the noise) and each of several Track 2 segments that either include the utterance sounding like "play" or lie on either side of it. (Specifically, we analyzed segments located at $t_{2\text{pm}} = 4:58.098$ to 4:59.618, 4:59.331 to 4:59.618, 4:59.618 to 4:59.988, and 5:00.017 to 5:00.371. The third of these segments contains the word that sounded like "play." Each Track 2 segment was speed-warped by the factor K_2/K_7, in order to convert Track 2 playback time intervals into equivalent Track 7 playback time intervals, before performing the cross-correlation.) Each of the resulting cross-correlation plots shows a clear peak at a sharply defined time. These timings are used to determine what point on Track 7 corresponds to the starting point of each Track 2 segment. For example, we find that the beginning of the fourth segment on Track 2 (immediately following the segment containing "play") corresponds to Track 7 at 15:56.707. We also find that between the second and third segments of Track 2, the corresponding point on Track 7 is delayed by 3.602 s (in playback time) compared to where it would be if there were no groove jumps on playback of the Gray Audograph disk. This delay is not the result of a skipback (repeat) during Track 7 recording, since such a repeat would cause additional cross-correlation peaks that are not present. It is instead the result of a groove skipforward during Track 2 recording, by exactly one rotation of the Audograph disk.

In reality, "play" is not present on Track 2, which accounts for it not being on Track 7, either. What is on Track 2 is "ay" that begins abruptly as the playback stylus of the Audograph jumped ahead one groove. On Track 7 one hears clearly, ". . . Dispatcher on One seems to be have his mike stuck. [screech] Get the trucks out of the way [period of quiet]."

On Track 2 one hears even more clearly the passage from "Dispatcher" through the period of quiet, except that a portion starting near the end of the "screech" and ending with the "w" of "way" has been elided. When we elide the same portion from Track 7, it sounds just like Track 2. In this region reproducing Track 7 at 8820 samples/s corresponds to $8820/1.08 = 8170$ samples/s for Track 2. The corresponding spectrograms for these portions of Tracks 7 and 2 are shown in Figure 3.

Figure 3 Graphical evidence relating to the "Play" utterance (see text). For the spectrograms, the x axis denotes frame number; at 8820 samples/sec, each frame is 512 samples long and is stepped (advanced) 64 samples from the previous one; so each x unit corresponds to a time interval of one frame step, namely $64/8820 = 0.007256$ s. The y axis denotes the frequency band; each y unit corresponds to a frequency interval of $4410/256 = 17.23$ Hz. Top to bottom [(a)–(d)]: (a) Spectrogram of Track 7 from $t_{7\text{pm}} = 15:52.27$ to 15:53.43. Note the region $0 < x < 70$. (b) Spectrogram of Track 7 from $t_{7\text{pm}} = 15:55.87$ to 15:57.03, precisely 3.60 s after part (a). Note the region $70 < x < 150$. (c) Spectrogram of Track 2 from $t_{2\text{pm}} = 4:59.22$ to 5:00.29, showing the elision of one disk groove (3.60 s) of $t_{7\text{pm}}$, so that "Get the trucks out of the way." becomes ". . . ay." The scale of this figure has been stretched by a factor 1.08 to match the local speed ratio between Track 2 and Track 7. (d) Pattern cross correlation of Track 2 from $t_{2\text{pm}} = 4:59.22$ to 5:00.46 (of which the first 1.07 s is shown in (c)) against Track 7 from $t_{7\text{pm}} = 15:42.80$ to 16:05.25. The peak at $x = 4404$ is due to the $0 < x < 70$ region of part (c); that at $x = 4900$ is due to the $x > 70$ region. The separation between the peaks is 3.599 s. (The x value denotes PCC time shift measured in frame steps of 0.007256 s each.)



The utterance "15 car 2 . . . now . . . on Main": There is one clear case of a defect in Track 7 recording that occurs prior to GO. On Track 2, starting at $t_{2\text{pm}} = 1:13.3$, there are four utterances of "15 car 2." Listening and cepstral analysis confirm that the second and third of these utterances are identical, indicating a skipback (repeat) on Track 2. The first has different prosody from the second, and the fourth is by a different speaker. These utterances are followed by the phrase "now . . . on Main . . . probably just past Lamar." In contrast, on Track 7 starting at 11:43.1, there is a single full-amplitude utterance of "15 car 2," then a 7.9 s period from 11:44.9 to 11:52.8 during which there is no full-amplitude sound, but attenuated utterances including "15 car 2" and "now . . . on Mai" are heard. The final "n" of "Main" is then immediately heard at full amplitude at 11:52.8. It appears that although the sound level of Track 7 was reduced during this period, there is no evidence of a skip on Track 7 at this point, either forward or backward. (To avoid confusion, note that these "15 car 2" utterances starting at $t_{2\text{pm}} = 1:13.3$ are unrelated to, and occur long before, the Table 1 entry marked "First '15 car 2"' which occurs at $t_{3\text{pm}} = 0:03.9$.)

Another defect on Track 7 affecting the amplitude and slightly distorting the sound begins at 15:46 and continues to the end of the track. It is present on Track 6 as well, and seems to be on the tape from which we prepared those tracks. There is no effect on the timing and little effect on intelligibility. The reduction in amplitude is asymmetric; when the instantaneous value of the voltage corresponding to the acoustic signal is positive, the amplitude is reduced by a factor of about four; when the instantaneous value is negative, the amplitude is reduced by a factor of about 12.

The utterance "Stand by": Thomas has claimed (personal communication, 2002) that another utterance "stand by" (heard by us either as "stand by" or "and uh") occurs at 1:54.7 on Track 2 (also prior to GO), and has no counterpart on Track 7. However, listening to Track 2 from 1:42 to 1:58 and Track 7 from 12:13 to 12:28 we find evidence for two skipbacks (repeats) and a skipforward on Track 2, but for no skips on Track 7. Specifically, we transcribe Track 2 as: "1, {5(?), 5, 4. What traffic personnel do you have on}, {REPEAT}, {REPEAT}, Cedar Springs in the vicinity of the Field here? (noise) {whisper = Cedar Springs and(?)} And uh (noise), {brief whisper = ??} – INFERRED SKIP – Cedar Springs and(?) Mockingbird? (noise)." The notation "{REPEAT}" indicates that the first phrase in braces is repeated twice in immediate succession. The "inferred skip" is discussed below. "Whisper" refers to a low-amplitude utterance, and question marks refer to utterances that are not heard clearly. We transcribe Track 7 as "1,5(?), 5, 4. What traffic personnel do you have on Cedar Springs in the vicinity of the Field here? {whisper = Cedar Springs and uh} And uh {whisper = ??} former(?) on Cedar Springs and uh {whisper = ??} Cedar Springs and Mockingbird?" The phrase marked "{whisper = Cedar Springs and uh}" is followed by the same utterance (but now at full amplitude, rather than a "whisper") after 3.60 s of Track 7 playback time, which equals one turntable rotation. The "whisper" is one of many examples of "premonitory whisper repeats" that we have found on the recordings derived from the Gray Audograph disk.

The Track 2 full-amplitude utterance transcribed as "and uh" that immediately follows the whisper "Cedar Springs and(?)" is the phrase that Thomas refers to as "stand by." This utterance (whatever the two words actually are) does appear on Track 7: it is the first full-amplitude "and uh" of "And uh . . . former(?) on Cedar Springs and uh . . . Cedar Springs and Mockingbird?" Note also that the Track 7 phrase "former(?) on Cedar Springs and uh {whisper = ??}" is absent from Bowles, arguing for a Track 2 skip at the position marked "INFERRED SKIP" above.

In addition to these findings on direct listening, the recording time interval (after K-factor speed correction) between the two words (transcribed either as "stand by" or as "and uh") on Track 2 is found to be equal to that between the two words on Track 7 that we claim correspond to this utterance.

In summary, we conclude that this portion of Track 2 has two track repeats followed by a forward skip. Apart from these, every utterance on Track 2 in this interval matches an utterance on Track 7.

Our analysis of the above cases, in which a phrase present on Track 2 is claimed to be absent on Track 7, has shown no evidence for Track 7 skips. [Note that Thomas (personal communication, 2005) states that his assertion concerning skips was made regarding the FBI phonographic playback, and was not limited to the portion of that playback that constitutes Track 7.] Even if a forward skip were documented in the interval between GO and YOU, it would increase rather than decrease the inferred recording time interval between these utterances, thereby increasing the time interval by which "Go to the hospital" precedes the alleged gunshot sounds. However, if there were a documented forward skip, it would affect the integrity of Track 7 timing calculations, and would increase the importance of determining whether there might also be repeats (skipbacks) on Track 7 in the interval of interest (which would place GO closer to the alleged shots).

Independent of the question of possible forward skips, we have analyzed the region of interest for possible Track 7 repeats. By direct listening, we find that no audible utterance is repeated at anywhere near full amplitude within this interval. (However, the "premonitory whisper repeats" discussed below in connection with the cepstral analysis, in which an attenuated version of a sound is heard one Gray Audograph rotation time prior to the full-amplitude sound, are present throughout most of the interval.) A repeat could be missed on direct listening if it occurred either during a quiet interval or during a time when noise, rather than intelligible signal, was present. However, there are no intervals of either quiet or noise longer than 4 s between GO and YOU. A repeat consisting of a single skipback of one Gray Audograph disk rotation would occupy 3.6 s of Track 7 playback time for the first rotation, plus 3.6 s for the repeat, for a total of 7.2 s. Such a repeat (if present) would therefore have been found by direct listening. Note that this argument does not rule out the possibility that two or more skips (forward and/or back) in rapid sequence could in principle go undetected by direct listening. (Since Track 7 was derived from the playback of the Gray Audograph disk on a standard turntable, there is no mechanism requiring the net number of forward and backward skips on Track 7 to be approximately equal at any given time. This contrasts with the case for Track 2, which was derived from the playback of the disk on a Gray Audograph machine.) Therefore, mathematical techniques including cepstral analysis and auto- (and cross-) correlation are of particular value for Track 7 and were employed.

As noted above, a repeat caused by the turntable stylus jumping backward one revolution during creation of Track 7 would cause a cepstral peak at a time shift of approximately 15,876 samples. We performed cepstral analysis using a sequence of 95%-overlapping frames, each of length 40,000 samples (about 9.07 s). The results shows that most intervals do contain a weak cepstral peak at about $15,886 \pm 6$ samples. We attribute it to a "premonitory whisper repeat" phenomenon, in which the distortion of a groove on the Audograph recording is caused by the embossing of the following groove. Where sufficiently clear speech is present this "whisper" is quite audible. The cepstral analysis reveals the "whisper" even when there is noise on the recording rather than intelligible speech. A skipback, however, would correspond to a repeated full-amplitude signal, not to an attenuated whisper.

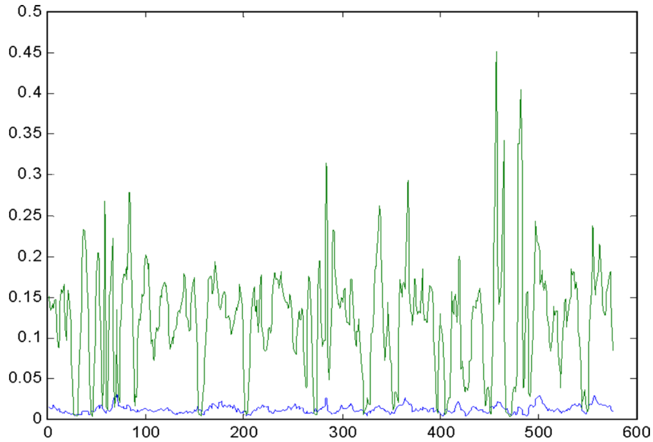
Specifically, we surveyed Track 7 from just before CHECK ($t_{7\text{pm}} = 12:39$) to just after YOU (17:00). The signal was digitally lowpass-filtered to 2 KHz, then downsampled to 4410 samples/s; each cepstral frame was 40,000 samples long, stepping each frame by 2000 samples; for a total of 576 frames. This was done both for the "natural" signal, and for an "artificial copy" signal in which samples #25,001 through 35,000 of each frame were copied at full amplitude to an interval 15,000 samples earlier, to replace samples #10,001 through 20,000.

Figure 4 shows, for each frame, the maximum value of the magnitude of the cepstrum in the vicinity of (i.e., from 20 samples below to 20 samples above) the expected time shift, which is about 3.6 s or 15,876 samples for the "natural" signal (lower curve), and 15,000 samples for the "artificial copy" signal (upper curve). We find that the "artificial copy" cepstral maximum is typically about ten times as large as the "natural" signal maximum for the same frame, showing that (for these frames) the natural signal contains no repeat at or near full amplitude and having a duration of the order of a second or more. In "quiet" frames, containing a low level of acoustic activity, both the "natural" and "artificial copy" cepstral maxima have a small value as expected, and this value is comparable to the background cepstral value (i.e., there is no clear cepstral peak in those cases).

These results show that, during the interval from CHECK to YOU, there is no evidence of a Track 7 repeat, and furthermore that the set of subintervals within which a repeat could be "hidden" is small. It appears extremely unlikely that 30 s of repeats would be "hidden," by chance, within this small set of sufficiently "quiet" subintervals. Furthermore, as Figure 1 shows, a repeat of 30 s on Track 7, if it were present, would correspond to a striking change in K by an amount 0.0125, which is not observed.

Although the variable playback speed for the FBI copy (Track 7) slightly complicates the calculation of recording times from playback observations, it adds greatly to the value of Track 7 in that it makes possible unique determinations of the AC hum

Figure 4 Cepstral survey of Track 7 including interval from CHECK to YOU utterances. Abscissa is the frame number; ordinate is magnitude of the cepstral maximum for that frame. See text ("Results: Analysis of Tracks 7 and 2 timing . . .") for details.



and hence the speed correction factor K. For this reason Track 7 is extensively used in the present paper. The constant interval per groove provided by the turntable playback that created Track 7 also provides a unique offset time for detecting repeats and simplifies cepstral analysis.

Correcting for the tape break between tracks 2 and 3 (Bowles copy of Channel 2)

To compute the Channel 2 recording times using the Track 2 and 3 (Bowles) playback times, we use (a) the K factors derived above; (b) the inferred time interval between the end of Track 2 and the beginning of Track 3 (owing to a tape break); and (c) the fact that the net effect of Gray Audograph stylus skips and repeats on the Bowles playback is small (on the order of 8 s or less) owing to the mechanical linkage that forces stylus skips in either direction to be compensated by skips in the reverse direction.

Using K_2 as derived above, we have calculated playback and recording times for the Track 2 recorded phrases up to the tape break as listed in the upper part of columns D and E of Table 1, on the assumption that the accumulated skips and repeats are balanced at the time the playback times were read. As can be seen from Table 1, the recording time between GO and the word "Dispatcher" on Track 7 is $67.6 + 78.3 = 145.9$ s, and on Track 2 it is $63.9 + 83.5 = 147.4$ s, in good agreement. However, to get times beyond the break, adjustments must be made for recordings lost in the break, for a new recording start up time, and for a different K. This is done as follows.

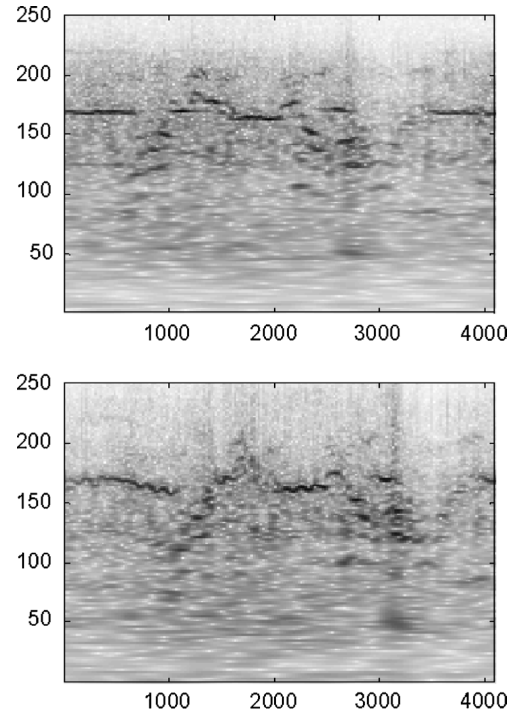
From column E of Table 1, $t_{2\text{rs}}$ at the tape break is 110.3 s. There are two ways we can obtain the time lost on the tape break. The NRC report ([3], p. 61) quotes Barger as saying that 0.4 s was lost in the break. We have measured a 1.0 s start up time at the beginning of Track 3 after the timing starts but before recorded sounds begin, so to get times that continue smoothly from Track 2 we need to add to the Track 3 reading $110.3 + 0.4 - 1.0 = 109.7$ s. Alternatively the last clear phrase before the break is "Dispatcher" and the first one after is "15 car 2." The recording time interval between these phrases on Track 7 is $113.4 - 78.3 = 35.1$ s. The recording time on Track 2 from "Dispatcher" to the tape break is $110.3 - 83.5 = 26.8$ s and the recording time on Track 3 from the tape break to "15 car 2" is 3.9 s. The net recording time that was lost is therefore $35.1 - 26.8 - 3.9 = 4.4$ s. So, to obtain the equivalent of Track 2 recording time from Track 3 one must add to the Track 3 recording time the sum of the Track 2 recording time at the time of the tape break and 4.4 s. Thus $110.3 + 4.4 = 114.7$ s must be added to the Track 3 recording time. Since $K_3 = 1.01$, the effective times after the tape break are given by $(114.7 + t_{3\text{rs}}) = (114.7 + 1.01 \times t_{3ps})$; this expression is used to obtain the Track 3 quantities in Column E (enclosed in curly brackets). We favor using this procedure to correct for our tape break, since it is based on the actual recordings that we used. In any case, Barger's 0.4 s loss alternative can be obtained from the one we are using by subtracting $(114.7 - 109.7 =)5.0$ s from the GO to YOU times that we calculate.

As previously stated, the numbers in the above paragraph and in Column E of Table 1 are based on the assumption that the accumulated skips and repeats are balanced. However this is not necessarily the case and there is uncertainty as to the number of unbalanced repeats, though the Gray Audograph mechanism makes it unlikely there would be more than one or two unbalanced repeats. Since the length of each studied repeat is less than 4 s and only about a quarter of the studied repeats are double repeats ([3], p. 63) the duration of any interval determined from Column E should be uncertain by less than 8 s. This renders the Bowles tapes imprecise for measuring short intervals of time, but more precise in percentage terms for longer time intervals. Note in particular that, in Table 1, the recording time for GO to YOU is $63.9 + 148.4 = 212.3$ s on Track 2, compared with the same time interval measured on Track 7 which is $67.6 + 143.2 = 210.8$ s, a difference of less than 1%.

Finding of warble or flutter on track 5 (FBI copy of Channel 1)
In Figure 5 we compare the "Gabor spectrograms" (defined in "Methods" section above) that include the first part of the YOU crosstalk utterance, for Tracks 1 (Figure 5a) and 5 (Figure 5b) (both Channel 1). Note the horizontal line segments found between bands #163 and 170 (i.e., between approximately 1550 and 1700 Hz) in the Track 1 plot. Corresponding to this in the Track 5 plot is a quite irregular line, showing that the Track 5 recording was made with considerable "warble" or flutter. Our analysis of this implies that the recording of Track 5 was subject to an irregular speed variation with amplitude of ± 3%, at a frequency of approximately 20 Hz. This flutter is found on other parts of Track 5 as well (including the vicinity of the HOLD utterance).

This finding was presented to Bruce Koenig, who supervised the recording of the NRC Committee copies produced at the FBI in 1981. According to Koenig (personal communication, 2003): "Since the Dictabelt loop had shrunk in size over the ensuing years, a common manifestation of this type of media, [William Sturtevant of the Dictaphone Corporation] had trouble inserting and playing the media on the playback device. He made various mechanical adjustments, including bending metal components on the unit, to allow playback of the Dictabelt. I believe that the media shrinkage and the mechanical changes probably produced the speed variances in question."

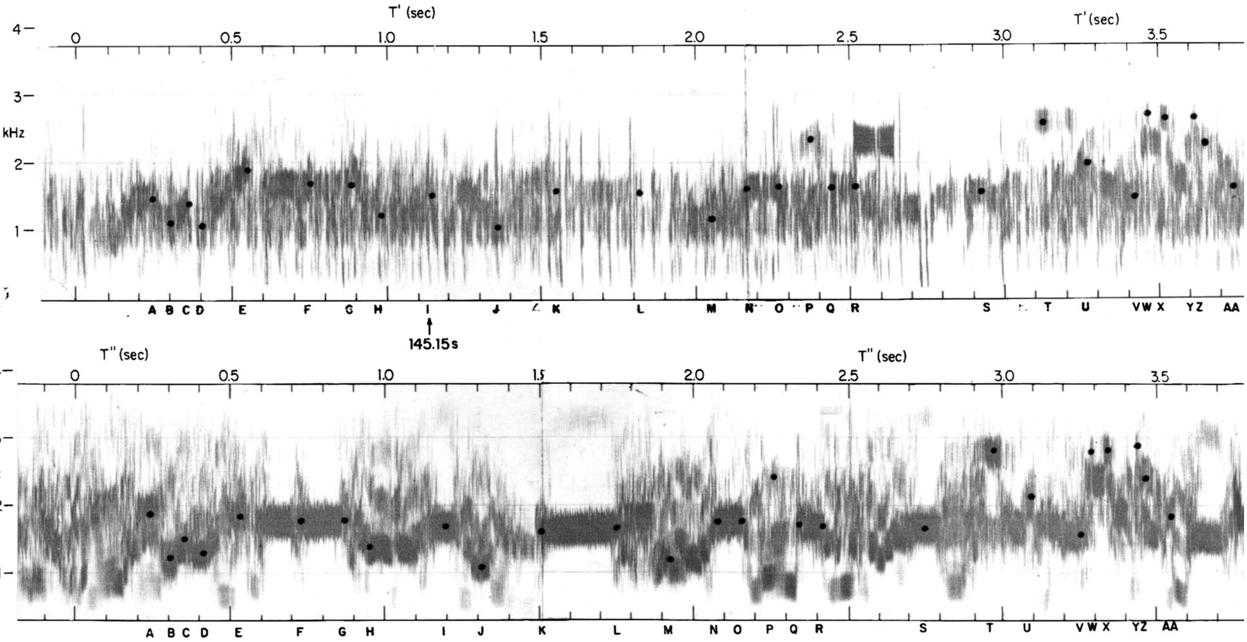
Figure 5 "Gabor spectrograms" [10] of the beginning of the YOU utterance on Track 1 (top) and Track 5 (bottom), showing Track 5 flutter. See text ("Methods: 'Gabor spectrogram' analysis" for details.) Abscissa denotes time [each x unit = (8 s/22050) = 0.3628 ms; full scale = 1.486 s]. Ordinate is filter band number (50 bands per octave), spaced evenly in the logarithm of frequency. The filters' center frequencies span the 5-octave range from 160 Hz (referred to as band #1) to 5120 Hz (band #250). Within each plot, the darkness of the image is proportional to the cube root of the absolute value of the Gabor filter output (to improve visibility over a larger range of output values).



Despite the rapid fluctuations in Track 5 speed at about 20 oscillations/s, the value of $K5 = 0.99 \pm 0.01$ averaged over time intervals greater than one second, as inferred above, is unaffected by this warble.

Timing of key acoustic events
GO on Channel 2 occurs at recorded times $t_{7\text{rs}} = -67.6$ and $t_{2\text{rs}} = -63.9$. There are two different utterances of "Go to the hospital": one at $t_{7\text{pm}} = 13:13.1$ (henceforth referred to as "GO") and a

Figure 6 Sound spectrograms ("voiceprints") of Channels 1 and 2 reproduced from the original photographs that provided the poor quality copies in Figure B-3 of the NRC Report. HOLD begins at approximately zero on both channels. The alphabetic notations and the dots are not referred to in this paper. [Reprinted with permission from "Report of the Committee on Ballistic Acoustics" by the National Academy of Sciences, courtesy of the National Academies Press, Washington, DC.]



second at 13:18.7 (referred to here as "Loud GO," but probably the utterance called "GO" in the NRC report).

To determine the claimed timings of the alleged first shot on Channel 1, we refer to BRSW. Note that the time that BRSW denotes as zero is different from our time origin, so it is necessary to derive the value of this time offset. As discussed in the NRC report, this can be most accurately done by comparing the sound spectrum in the upper part of Figure 6 with the sound spectrum in the BRSW report ([1], p. 26, Figure 5). This comparison places the alleged third or Grassy Knoll shot at the time shown by the arrow in Figure 6, with 145.15 being the time on the scale used by BRSW. The nearest clear word to this is HOLD on Tracks 1 and 5, which is at 0.0 on the spectrogram at $t_{5pm} = 3:16.1$ and at $t_{1pm} = 3:57.4$. The alleged first shot according to BRSW ([1], p. 10) occurs 7.6 s before the Grassy Knoll shot, which in turn is 1.13 s after HOLD in Channel 1 playback time, so the alleged first shot should be at $t_{1pm} = 3:57.4 - 7.6 + 1.13 = 3:50.9$, as given in Table 1.

Analysis of putative crosstalks: YOU, HOLD, and CHECK

For each of the putative crosstalks, we give (a) the locations of the intervals containing the utterances on Tracks 1 and 7; (b) the optimum value of the "warp" (speedup factor applied to Track 1) that maximizes the peak of the pattern cross-correlation (PCC) function; and the value of that PCC peak (as well as the approximate value of the background PCC value in the vicinity of the peak); and (c) the optimum value of the additional unphysical "duration-only" warp ("d-warp," a mathematical warping of time that keeps frequencies unchanged). Figures 7–9 present illustrative PCC plots and spectrograms for each putative crosstalk.

YOU (Figure 7): We used Track 1 from $t_{1pm} = 6:49.363$ to 6:51.958 (an interval 2.595 s long); and Track 7 from 16:44.982 to 16:57.077 (12.095 s long). The optimal speed warp was 1.012 (i.e., corresponding to a 1.2% speedup of Track 7). For this warp, the PCC peak value is 0.39, compared with a neighboring background value of approximately 0.10 (Figure 7c). Holding the speed warp at 1.012, and varying the "d-warp" from 0.96 to 1.05, the value of the PCC peak minus background increased from 0.19, to a maximum of 0.28 at a "d-warp" factor of 1.00 (corresponding to zero non-physical time stretch), then decreased to 0.16.

HOLD (Figure 8): We used Track 1 from 3:57.481 to 3:59.975 (2.494 s long), and Track 7 from 14:19.993 to 14:35.246 (15.253 s long). The optimal speed warp was 1.055, corresponding to a 5.5% speedup of Track 7. For this warp, the PCC peak is 0.32, compared with a neighboring background value of approximately 0.13 (Figure 8c). Holding the speed warp at 1.055, and varying the "d-warp" from 0.97 to 1.05, the value of the PCC peak minus background increased from 0.12, to a maximum of 0.20 at a warp factor of 1.00 (again corresponding to zero non-physical time stretch), then decreased to 0.10.

CHECK (Figure 9): An utterance "I'll check it" (here denoted CHECK) occurs on Channel 2 (at 2:07 on Track 2 and at 12:39 on Track 7). It has been claimed by Thomas (personal communication) and others that this utterance also appears as crosstalk on Channel 1 (at 3:45 on Track 1 and at 3:05 on Track 5), and that the

Figure 7 Spectrograms and pattern cross-correlation (PCC) plots for putative crosstalk "YOU." Axes and units are as defined in Figure 3. From top to bottom [(a)–(c)]: (a) Spectrogram of Track 1 from $t_{1pm} = 6:49.36$ to 6:51.96, stretched by a factor 1.012 for best local speed match to Track 7. (b) Spectrogram of Track 7 from $t_{7pm} = 16:51.08$ to 16:53.68. (c) PCC of "YOU want me to still hold this traffic on Stemmons" utterance, for Track 1 ($t_{1pm} = 6:49.36$ to 6:51.96) against Track 7 ($t_{7pm} = 16:44.98$ to 16:57.08), plotted for best local stretch of Track 1 by a factor 1.012. The peak at $x = 2360$ is located at $t_{7pm} = 16:51.08$.

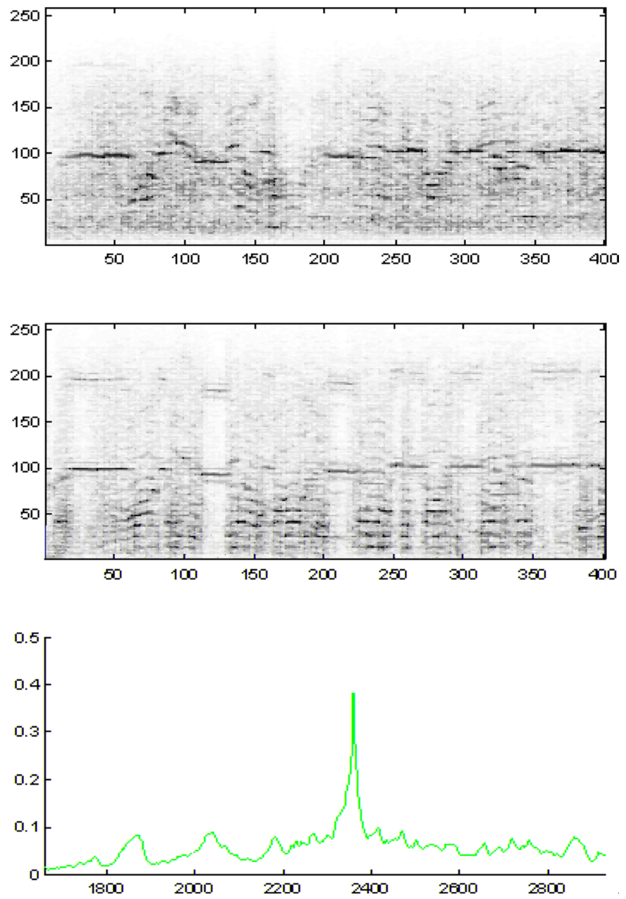
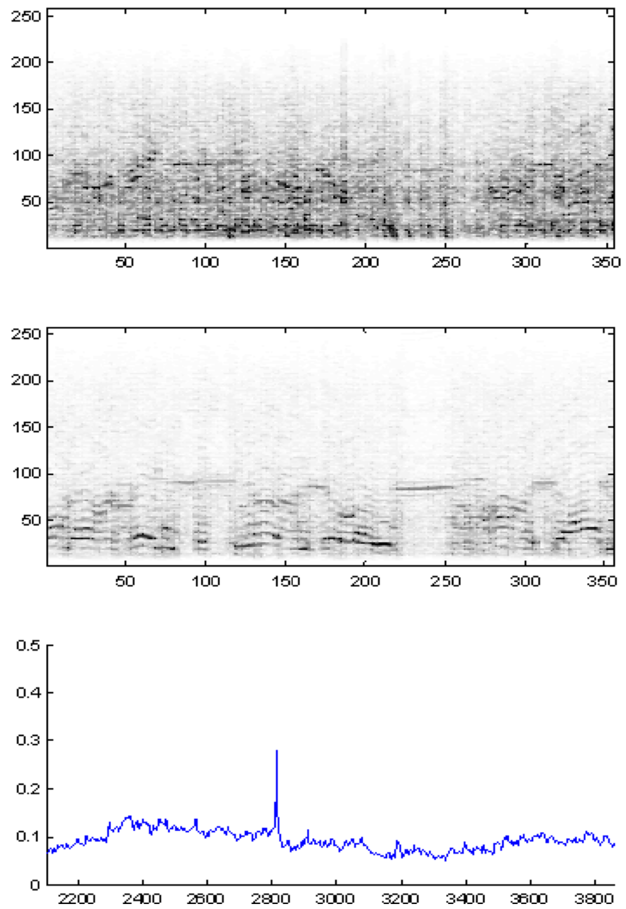


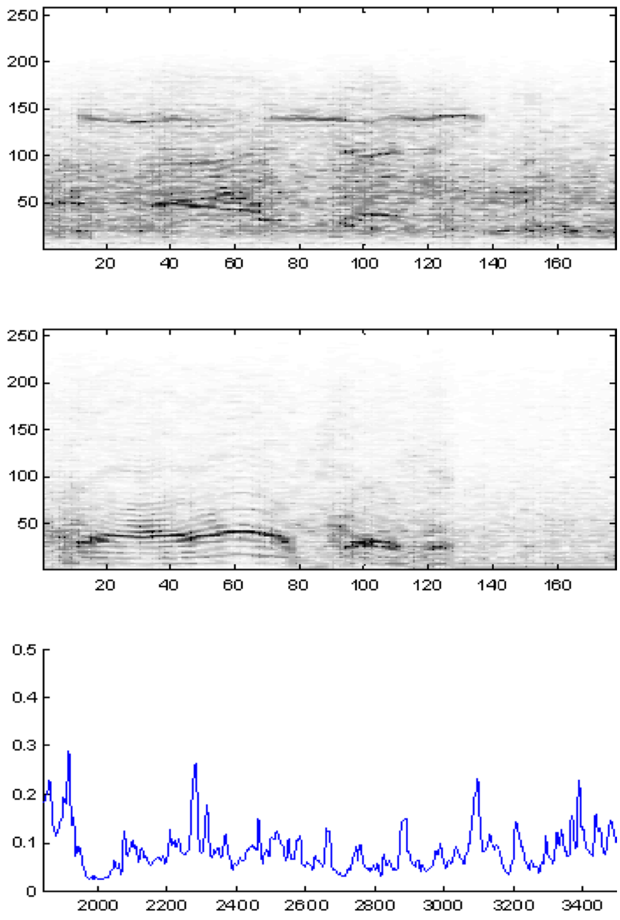
Figure 8 Same as Figure 7 but for putative crosstalk "HOLD." (a) Spectrogram of Track 1 from $t_{1pm} = 3:57.481$ to 3:59.975, stretched by a factor 1.055 for best local speed match to Track 7. (b) Spectrogram of Track 7 from $t_{7pm} = 14:25.21$ to 14:27.76 ("HOLD everything secure until Homicide"). (c) PCC of "HOLD everything secure until . . . " utterance, for Track 1 from $t_{1pm} = 3:57.481$ to 3:59.975 (2.494 s duration), against Track 7 from $t_{7pm} = 14:19.993$ to 14:35.246 (15.253 s duration), plotted for best local stretch of Track 1 by a factor 1.055. The peak at $x = 2814$ is located at $t_{7pm} = 14:25.21$.



timings of this putative crosstalk are incompatible with HOLD. The very noisy Channel 1 utterance has been heard by various listeners as "I'll check it," as "I'll get it," or as other quite different words. In Table 1 we denote this Track 1 utterance as "CHECK1." If CHECK were a valid crosstalk, its timing would be incompatible not only with HOLD, but also with the timing of the well established crosstalk YOU, as can be seen from Table 1 and the following argument. From Column C the Track 7 recording time from CHECK to YOU is $99.0+143.2$ s $= 242.2$ s. However, from Column G the Track 1 recording time from CHECK1 to YOU is $12.4 + 173.0 = 185.4$ s. This discrepancy cannot be blamed on unrecorded Channel 1 dead time on Track 1, since the motorcycle microphone was stuck open during this time.

To analyze this putative crosstalk, we used Track 1 from 3:46.286 to 3:47.635 (1.349 s long), and Track 7 from 12:33.454 to 12:46.822 (13.369 s long). The audio for "I'll check it" on Track 7 runs from about 12:39.23 to 12:39.96, beginning about 5.80 s after the start of the Track 7 segment. The audio for "I'll check it" on Track 1 begins about 0.28 s after the start of the Track 1 segment. The PCC method gives a peak at the expected delay that is no larger than other PCC peaks that wax and wane as the speed

Figure 9 Same as Figure 7 but for putative crosstalk "CHECK." (a) Spectrogram of Track 1 "I'll check it" from $t_{1\,pm} = 3:46.286$ to 3:47.635 (duration of 1.349 s), stretched by a factor 1.08 to match the tape speed of Track 7 in this region. (b) Spectrogram of Track 7 "I'll check it" from $t_{7pm} = 12:38.896$ to 12:40.245 (duration of 1.349 s). (c) PCC of Track 1 "I'll check it" from $t_{1pm} = 3:46.286$ to 3:47.635 (1.349 s duration) against Track 7 from $t_{7pm} = 12:33.454$ to 12:46.822 (13.369 s duration). If "I'll check it" shown in part (a) were a valid crosstalk, a large correlation peak should appear at index $x = 2603$. (Note that the peak at $x = 2579$ is very small, and does not vary significantly with non-physical "d-warp.")



warp is varied from 0.85 to 1.14. Figure 9c shows a typical PCC plot, at the speed warp of 1.08 that corresponds to the speed ratio appropriate to the values of K_1 and K_7. (The PCC peak for the similar features in Figure 9a and b should appear at a delay near $x = 2603$. The peak at $x = 3130$ is 3.5 s away and has nothing to do with "I'll check it" on Track 7.) This lack of a prominent peak would be expected if the two instances of "I'll check it" were different utterances on the two channels and not a crosstalk. In addition, it is quite evident that the prominent frequencies of Figure 9b are lower than the corresponding frequencies of Figure 9a, despite the frequencies having been lowered by 8% by the time stretch in Figure 9a. Arbitrarily adjusting the speed ratio to match the frequencies would bring the intra-utterance timings out of alignment. The peak at $x = 2663$ is maximized at a physical speed warp of 1.12 and has an amplitude of only 0.07 above its neighboring background. This peak increases to amplitude 0.10 when a non-physical "d-warp" of 0.80 is combined with the physical speed warp of 1.12. The peak at $x = 2579$ has amplitude of only 0.05 above its neighboring background and does not change in amplitude as non-physical "d-warp" is varied from 0.80 to 1.18.

For comparison, we studied some clearly repeated utterances (i.e., the same words uttered twice, apparently by the same person) on a single channel, such as "Ten–four." A pair of occurrences of this utterance on Channel 1, within a few seconds of each other, were analyzed by the PCC technique. The optimum physical warp was 1.11 (rather than 1.00), reflecting a difference in speech and not in tape speed, and there was only 3% variation in the peak amplitude of 0.70 as this optimum warp of 1.11 was combined with a "d-warp" that was varied from 0.90 to 1.08.

As discussed in this section and in the Methods section earlier, the PCC determines the relative timing and speed of the two channels, and can confirm the presence of putative crosstalk. In the case of YOU and HOLD, our findings show strong evidence of crosstalk: when the physical speed warp is chosen to maximize the PCC peak, the optimal "d-warp" then has a value of 1.00 (i.e., no non-physical stretch of the time axis). However, when comparing CHECK with its putative crosstalk CHECK1, the PCC peak near the appropriate delay that matches the perceived position of the two utterances is smaller than many clearly accidental peaks, is maximized at a warp departing by 4% from that appropriate to the tape speed ratio (i.e., 1.12 versus 1.08), and is not sensitive to variation of the "d-warp."

Even if the same words "I'll check it" appear on both channels, we conclude that they were spoken separately, and at different times.

Synchronization of the two channels, and time from GO to first alleged shot

Suppose we synchronize using the crosstalk HOLD. This utterance occurs at $t_{7rs} = 0.1$ and $t_{2rs} = 0.4$ on Channel 2, and at $t_{5rs} = 194.1$ and $t_{1rs} = 0.1$ on Channel 1. Therefore the interval of recording time from GO to HOLD is 67.7 s for Track 7 (FBI), and $63.9 + 0.4 = 64.3$ s for Track 2 (Bowles). Since (as discussed earlier) we find no skips or repeats on Track 7 during this period, and there was no dead time during this period, the actual elapsed time from GO to HOLD is equal to 67.7 s. Using the BRSW timing of the "first shot" (see above and Table 1) we calculate that the recording time corresponding to the "first shot" is $t_{1rs} = -6.5$ and the recording time from the "first shot" to HOLD is 6.6 s; this is the actual time elapsed, since Channel 1 had no dead time. Therefore, in actual time and using HOLD crosstalk synchronization, GO was recorded $67.7 - 6.6 = 61.1$ s before the alleged first shot.

By this analysis, the beginning of the phrase "Go to the hospital" therefore precedes the first "shot" by approximately 1 min.

In further support of the validity of HOLD being a valid instance of crosstalk, we note that there is also a bell-like tone (called BELL-b in Table 1) which in recording time occurs 7.7 s after HOLD on Track 7 (Channel 2) and 7.8 s after HOLD on Track 1 (Channel 1). Unfortunately the source of the bell tone is unknown; it might have been a continuation of the HOLD crosstalk, an electrical artifact, or a sound picked up in the DPD office. Whatever the source, this tone provides further support for the validity of HOLD being an instance of crosstalk, either by showing that the crosstalk lasted for at least 8 s or by providing an independent instance of crosstalk synchronization.

Suppose we instead synchronize the channels using the crosstalk YOU. This utterance occurs at $t_{7\mathrm{rs}} = 143.2$ and $t_{3\mathrm{rs}} = 148.4$, and at $t_{1\mathrm{rs}} = 173.0$. The Channel 2 recording time interval from GO to YOU (derived from Track 7 data) is therefore $67.6 + 143.2 = 210.8$ s; the same interval derived from Track 2 and Track 3 data is $63.9 + 148.4 = 212.3$ s. The Channel 1 recording time interval from the first "shot" to YOU (derived from Track 1 data) is $6.5 + 173.0 = 179.5$ s; the same interval derived using Track 5 is $368.7 - 187.7 = 181.0$ s. Combining these results yields an actual elapsed time from GO to the first "shot" equal to 31.3 ± 1.5 s plus the Channel 2 dead time during the interval from GO to YOU.

Regarding the dead time on Channel 2, note that the interval from GO to YOU contains five periods of radio silence lasting at least 4 s each, during any of which the recorder should have stopped, but for an unknown period of time.

The above results from HOLD and YOU synchronization—61 s and "31 s plus Channel 2 dead time," respectively—are consistent with one another, and imply that the dead time should equal about 30 s.

Discussion

The NRC report

The NRC report [3] relied primarily on the HOLD crosstalk data and the FBI recording corresponding to Track 7. We essentially confirm the NRC analysis using that data and we also conclude that the alleged first shot occurred one minute after "GO to the hospital." However, we have found several errors in that report.

We have followed the procedure developed by the NRC Committee to analyze the Track 7 tape, by using the YOU crosstalk for synchronization, but we obtain slightly different numerical results. Thus we find the time from GO to YOU from our data is 210.8 s whereas NRC [3] gives 206 s. We suspect that the NRC may have used the "Loud GO" utterance (defined above) instead of the GO we used, accounting for this difference of approximately 5 s. This difference does not affect any of our conclusions.

The NRC Committee made a significant error in analyzing the Bowles tapes. The Committee identified many repeats, some of which were immediately followed by a skip forward, but at that time did not realize that the relative positions of the stylus and the rotating recording disc of the Gray Audograph were mechanically driven, both during recording and playback, and so repeats should, on the average, be compensated by skips. As a result, in the analysis of this recording 18 s were incorrectly subtracted for repeats, making GO appear to be closer to HOLD and YOU. Therefore, almost all of these 18 s (at least about 14 s) should be added back in Table C-1 of the NRC report. Note that this correction acts to strengthen the NRC's conclusion. Another error, this one procedural, was made in determining the recording speed. Since Figure 6 (copied from the NRC Report) contains sound spectrograms of both Channel 1 and 2, the two channels were compared in both time and frequency leading to the conclusion that times on Channel 2 should by multiplied by a factor of 1.06 to agree with times measured on Channel 1. Thomas (personal communication, 2002) has correctly pointed out that the single sound spectrogram lasted for only 3 s, which does not justify applying this factor of 1.06 over the entire region of interest. In the present work we have found constant-frequency hums on Tracks 1, 2, 3, and 5, and precisely linearly-varying AC hum on Track 7, throughout the entire relevant portion of each track, and from these measurements and PCC matches have obtained the correction factor K for each of these tapes. (Note that these values yield $K_2/K_1 = 1.06$, consistent with the value used by the NRC Committee.) Since the NRC Report incorrectly subtracted time for repeats, Table 1 of the present paper should be used instead of the NRC Report's Table C-1. Since the NRC Committee primarily relied on the HOLD crosstalk, the YOU crosstalk, and the speed-calibrated FBI copy of Channel 2, the Committee's general conclusions were not significantly affected by the above errors made in the analysis using the Bowles tape.

The Thomas paper

Although Thomas [5] based his article in part on the analysis of the NRC report, he primarily used the portions pertaining to the Bowles copy of the Gray Audograph recording. The Bowles copy contains many repeats, and the NRC considered it to be less reliable than the FBI copy that was primarily used by the NRC. In analyzing this data, Thomas made the following errors.

(1) Thomas ([5], p. 29) states: "The NRC panel failed to recognize the synchronization that arises from using the Bellah cross-talk episode because instead of using real time to compare the two tapes they used artificial time, what they referred to as 'channel one' time. Because they used artificial time instead of real time, they failed to recognize the need to correct for the warp in tape speed." This statement is incorrect. The NRC panel not only recognized the synchronization from the Bellah crosstalk (i.e., the "YOU ... Stemmons" utterance), but devoted most of the report's 20-page Appendix C to that crosstalk. As discussed earlier in this paper, there are several different possible time scales that are equally valid if used consistently, as was done in the NRC report. Since the Channel 1 recorder had no interruptions, Channel 1 playback time can be converted directly to "actual time" by multiplying by the time correction factor K. Finally, the NRC panel not only recognized the warp in tape speed, but also corrected for it as appropriate.

(2) Thomas ([5], p. 29) states: "Thus if one uses the Bellah cross-talk to synchronize the transmissions of the two police channels, instead of the Decker calls, then the putative gunshots exactly overlap the interval of time defined by Chief Curry's two broadcasts and occur at the exact instant that John F Kennedy was assassinated." He reaches this conclusion by using NRC report Table C-1 entries that state that the Channel 2 playback time from GO to YOU with repeats subtracted is 180 s, whereas the Channel 1 playback time from the alleged shots to YOU is 171 s. He then assumes (in our notation) a speed correction factor of $K = 1.05$ for Channel 1. His argument also implicitly depends on assuming $K = 1.00$ for Channel 2. He therefore finds the Channel 1 recording time interval from the "shots" to YOU to be 179 s, in agreement with the Channel 2 recording time interval of 180 s from GO to YOU. However, there are three flaws in this analysis (see "Results" above). First, the NRC erred in subtracting the full 18 s for repeats and almost all of this time, say 14 s, should be added back in. Second, the implicit assumption that $K = 1.00$ is unwarranted. (One might expect, or implicitly use, a value of 1.00 because Bowles used the same Gray Audograph for playback and recording. But this would not be true if the Audograph speed during playback differed from what it was at the much earlier time of the original recording.) Third, we have determined that $K_1 = 1.01$, $K_2 = 1.07$, and $K_3 = 1.01$. With all these corrections, the recording time from GO to YOU on Channel 2 is (see Table 1 and results above) 210.8 s using Track 7, and 212.3 s using Tracks 2 and 3, whereas the recording time from the first alleged shots to YOU on Channel 1 is 179.5 s using Track 1, and 181.0 s using Track 5. The first alleged shots are thus placed at approximately 31 s, plus Channel 2 dead time, after the assassination.

(3) Thomas ([5], p. 29) states, with respect to the use of the dispatcher's time annotations for relating playback time to actual time, and to BRSW's regression analysis for computing a best linear fit of playback time vs. annotated time: "But over the six minutes immediately after Curry's broadcasts the slope of the regression line was a perfect 1.0. Thus there can be no significant amount of lost time on Channel 2 after 12:30 [P.M.] . . . ." This statement is based on BRSW's statement that their data indicated a least square error fit slope of 1.0. However, the data on which the BRSW statement is based are plotted in ([1], p. 32). The measured slope of the line in that plot that shows the regression of Channel 2 *playback* time (the *y* axis) against annotated time (the *x* axis) between 12:30 and 12:36 P.M. is 0.94, not 1.0. When we perform the same regression on the same six points used by BRSW, we likewise obtain a best-fit slope of 0.94 ± 0.05, where throughout this discussion the number following the "$\pm$" symbol denotes one standard deviation. (Note also that only six of the seven annotations were used in the BRSW plot; one of the "12:36 [P.M.]" annotations was omitted.) Also (since it is playback time, and not recording time, that is used by BRSW), even if the slope were unity, it would not follow that there was no dead time without also using the assumption that the ratio K of recording time to playback time is also unity (as Thomas does implicitly), and there is no basis for this extra assumption. Finally (see Results section), the two annotations for "12:35 [P.M.]" occur about 30 s apart, and the same is true for "12:36 [P.M.]." The slope derived from data having this degree of scatter is too imprecise to support a conclusion that there is an insignificant amount of dead time, as we show below.

The dispatcher's time annotations indeed provide an important test since any valid time calibration should be compatible with the time annotations, but the time annotations by themselves do not provide reliable time calibrations. As long as the possibility of dead times exist, one must make specific assumptions about the dead times to get a calibration. The assumption that the Channel 2 dead times between annotations are zero (or negligibly small) is just as specific an assumption as saying that they are, for example, 20 and 10 s, as discussed below. To favor one time calibration over another, the investigator must show that one regression analysis is better than the other to a statistically significant extent, and this was not done. We have shown that time scales with sufficient allowable dead times are compatible with the data.

To illustrate how various amounts of dead time are compatible with the annotations and the regression analysis, we perform a regression of Track 7 (Channel 2) *recording* time – as modified by adding the various amounts of dead time – against annotated time (on the *x* axis), using all seven annotations, and making various assumptions regarding the amount of dead time. If zero dead time is assumed between 12:30 and 12:36 P.M., the best-fit slope is 0.952 ± 0.050. If, instead, one were to assume that the Channel 2 recorder stopped for 20 s between 12:30 and 12:32 P.M. and for another 10 s between 12:32 and 12:34 P.M., then the slope of the regression curve would be 1.03 ± 0.05. Other assumptions also involving a total of about 30 s of dead time give similar results. Note that both of these slopes are consistent with a slope of 1.00. In fact, since the standard deviation is 0.05, any best-fit value lying between about 0.90 and 1.10 (that is, within two standard deviations of unity) would be statistically consistent with a slope of unity. (The standard deviation of the slope is so large because of the large scatter in the data points, as noted above.) With the added 30 s of dead time (as assumed for this illustration), the recording time plus dead time from HOLD to YOU would be $143.2 - 0.1 + 30 = 173.1$ s on Channel 2 (using Track 7), compared with 172.9 s on Channel 1 (using Track 1). This illustration does not prove that these two particular dead times are correct, any more than the BRSW data prove there is no dead time. However, it does show, contrary to Thomas's assertion, that dead times totaling as much as 30 s, or even a somewhat larger amount, are compatible with the dispatcher's annotations and lead to regression curves that are as good as those that use his assumption of an insignificant amount of dead time.

(4) Thomas ([5], pp. 29–30) uses his claim of no significant dead time [item (3) above] to argue that the HOLD utterance cannot be valid for synchronizing the two channels, stating: "Because the regression analysis [of the dispatcher's time annotations] shows that no time is missing from the relevant section of the Channel 2 tape, then the fragment from Sheriff Decker's broadcast is only explained by the overdub

hypothesis." ["Overdub" here means that the utterance on Channel 1 appears on the recording medium in the incorrect location, as "the result of the recording stylus jumping backward in its track."] We have shown above that, on the contrary, since the dispatcher's annotations are indeed consistent with various amounts of dead time (as shown above), Sheriff Decker's broadcast ("HOLD") can thus be explained as a normal properly located crosstalk similar to other established crosstalks such as YOU. We now show, furthermore, that Thomas' "overdub" argument, taken together with his assertion of the validity of the CHECK crosstalk, lead to an arithmetic contradiction.

We identify the following elements that are directly stated or implied by Thomas' scenario in which an "overdub" of HOLD occurs, and the alleged shots precede "Go to the hospital" in actual time: (a) The acoustic image of the HOLD utterance on Channel 1 is positioned earlier on the recording medium than it should have been, owing to a stylus skipback on Channel 1. The time of the utterance that would be inferred from its position is therefore earlier than the actual time of the HOLD utterance, by an amount we will refer to as "SB" seconds of actual time ("SB" denoting "skipback"). (b) YOU is a valid "time tie" for synchronizing the two channels. In addition: (c) Thomas (personal communication, 2002) and others (e.g., Bowles [11]) have claimed that CHECK (on Channel 2) and CHECK1 (on Channel 1) constitute a "time tie" – that is, their locations on the recording media correspond to the same actual time.

We use only time intervals (between events on the same track) as shown in Table 1, which we derived only from playback times and K values as computed above. We denote the Channel 2 dead time (if any) between CHECK and GO by "DTCG," and that between HOLD and YOU as "DTHY." The actual time corresponding to the distance between the "overdubbed" acoustic image of HOLD on Track 1 (Channel 1) and YOU on Track 1 is $173.0 - 0.1 = 172.9$ s. Therefore the actual time from the utterance HOLD to YOU is (172.9-SB) seconds. On Track 7 (Channel 2) the recording time interval from HOLD to YOU is $143.2 - 0.1 = 143.1$ s. Therefore the actual time from HOLD to YOU is (143.1 + DTHY) seconds. Equating these two expressions yields SB = 172.9 − 143.1 − DTHY = (29.8 − DTHY) seconds. That is, on this scenario SB must be *at most* 29.8 s (since DTHY cannot be less than zero).

Next, the Track 1 recording time from CHECK1 to the "overdubbed" acoustic image of HOLD is $12.4 + 0.1 = 12.5$ s. Therefore the actual time from CHECK1 to the HOLD utterance equals (12.5 + SB) seconds. On Track 7, the recording time from CHECK to HOLD is $99.0 + 0.1 = 99.1$ s. Therefore the actual time from CHECK to HOLD is (99.1 + DTCH) seconds. Equating these two expressions yields: SB = 99.1 − 12.5 + DTCH = (86.6 + DTCH) s. Therefore, SB must be *at least* 86.6 s.

Since the two conclusions regarding SB in the previous two paragraphs cannot both be true, and in fact contradict each other by almost a full minute, the elements (a)–(c) above, which constitute Thomas' "overdub" scenario combined with the assertion that CHECK is a valid "time tie," cannot all be correct.

> Our analysis shows instead that (a) CHECK and CHECK1 are not a valid "time tie"; (b) there is no known evidence for a HOLD skipback on Channel 1; and (c) YOU is indeed a valid "time tie."

We have rebutted both the argument based on dispatcher time annotations and that based on the "overdub" hypothesis. There is every reason to believe that HOLD is a valid crosstalk. Especially compelling is the observed suppression of some of the cross-talk tones by strong heterodynes, proving that the crosstalk sounds arrived at the recorder via the radio channel and were not recorded later. The NRC report ([5], pp. 81–88) gives a number of reasons, including the sound spectrogram in Figure 6, for favoring HOLD as a genuine crosstalk. The validity of the HOLD crosstalk is further supported by the existence of the Bell-b sounds in Tracks 7 and 2 at corresponding times (see "Results"). As we have shown, using the HOLD synchronization the sounds alleged to be the first shot were recorded approximately 61 s after "Go to the hospital."

We have also eliminated (see "Results: Track 7" above) the challenges to the integrity of the Track 7 recording posed by Thomas's assertions that two utterances [a word "twenty" (or "play") and "stand by"] each appear on Track 2 and not on Track 7 at a corresponding time. In each case we have identified a skip or repeat on Track 2, and not on Track 7, as being responsible for the discrepancy between the two tracks.

General remarks

There have been many misinterpretations of the NRC report and we feel we should discuss them briefly before giving our final conclusions. Some have claimed the NRC report proved there was no conspiracy and others have claimed the report failed to prove there was no conspiracy. Both of these claims are misleading. As a general statement, it is essentially impossible ever to establish the *absence* of a conspiracy (unless every possible conspirator had been under observation all of the time), whereas in some cases it is possible to establish the *existence* of a conspiracy. One of the reasons that conspiracy theories thrive is that many people are uncomfortable with uncertainty. Some have given the conspiracy claims full credence, and have dismissed the NRC report by saying that the Committee merely found no evidence for a conspiracy.

The NRC Report and the present paper do far more than merely come up with no evidence for a conspiracy. They show that the evidence presented for a high probability of a second gunman is invalid because the sounds alleged to be shots occurred long after the President had already been shot.

In this report we have not directly addressed Thomas's calculation of the likelihood that impulses on the recordings are from gunshots and that there is a gunshot from the Grassy Knoll. Rather we have shown that his assertion, that these impulses were simultaneous with the assassination, is incorrect.

Since interest in the Kennedy assassination remains considerable after 40 years, there is every reason to believe there will continue to be interest for many years to come. However, the primary acoustic evidence is recorded on a fragile plastic Dictabelt and an Audograph disk that may become unreadable in the future. In January, 2004, NARA worked with a contractor to re-record several Dictabelts from November 24 and November 22, but the crucial belt that was the source of Tracks 1 and 5 was in such poor condition that it could not be played by the equipment available. NARA plans to use a new "optical stylus" technology to scan the Dictabelt, perhaps in 2005.

Conclusions

We have re-analyzed the NRC report [3] and have studied the Thomas article [5]. We have found errors in both articles. We have for the first time determined a consistent set of speed correction factors K for all relevant tracks, and shown that criticisms of the integrity of Track 7 timing are unfounded. The Channel 2 recording time measurements using the Bowles copies (Tracks 2 and 3) are found to agree well with those using the FBI copy (Track 7).

Thomas ([5], pp. 29–30) concludes that "the putative gunshots . . . occur at the exact instant that John F. Kennedy was assassinated" and that ". . . Sheriff Decker's broadcast is only explained by the overdub hypothesis." We have identified specific errors that led to Thomas's incorrect conclusions.

We have described the errors in some of the NRC panel's assumptions and analyses, and noted that their final conclusions remain valid for two reasons. The NRC analyses were primarily based on the HOLD and YOU synchronizations using the FBI copy in Track 7. The NRC Committee made no substantive errors in these analyses. Our correction (in the present paper) of an error in the NRC Committee's analysis of the Bowles tapes fortuitously strengthens the NRC Committee's basic conclusions. Thomas, on the other hand, used neither the HOLD synchronization nor the Track 7 analyses, but did use the erroneous parts of the Committee's analysis of the Bowles recordings and combined it with an erroneous implicit assumption that for Track 2 the time correction factor $K = 1.00$. These errors were in the same direction, contributing to his incorrect conclusion. Our present studies not only affirm the NRC Panel's conclusion but make an even stronger case by: correcting errors; removing the two objective criticisms to the HOLD crosstalk (i.e., those based on the dispatcher's time annotations and on the timing incompatibility of a HOLD crosstalk with a putative CHECK crosstalk); and showing that the analyses of Channel 2 based on the FBI copy (Track 7) agree with those based on the Bowles copy (Tracks 2 and 3), provided that the proper time correction factors K obtained using AC hum and spectrographic pattern cross-correlation are applied.

We affirm the NRC conclusion "that the impulses attributed to gunshots were recorded about one minute after the President had been shot and the motorcade had been instructed to go to the hospital." We also show that if, instead, the HOLD synchronization is ignored and the "YOU . . . Stemmons" synchronization is used, the first sounds alleged to be from shots occur at least 30 s after the assassination.

Most fundamentally, as emphasized in the NRC report, once one has established "Hold everything secure . . . " as a valid crosstalk in its proper position (no "overdub" involving a skipback on the Channel 1 recording), then no further timing analysis is needed to show that the impulses were not the assassination shots, because they overlap in time with "Hold . . . ," which by its meaning could only have been uttered after the assassination was recognized.

Acknowledgments

We thank Steve Barber, who played a crucial role in the NRC report by pointing out the crosstalk HOLD, and notified us of the independent analysis by Michael O'Dell, who has kindly shared his analysis. O'Dell independently made the key observation that repeats in the Bowles copy of Channel 2 must be largely compensated by forward skips; he also provided other information and insights in regard to the Audograph. We thank Professor Kenneth Stevens for sound spectrograms of the section from HOLD to YOU of Tracks 1 and 7. Micha Schattner did the transfers from reel-to-reel tape to digitized audio, and Chip Coldwell converted these to MP3.

References

1 Barger JE, Robinson SP, Schmidt EG, and Wolf JJ. Analysis of Recorded Sounds Relating to the Assassination of President John F. Kennedy (Appendix to Hearings Before the Select Committee on Assassinations of the House of Representatives Ninety-Fifth Congress, Volume VIII). Washington: U.S. Government Printing Office, 1979 (hereinafter referenced HSCA).

2 Weiss MR and Aschkenasy E. An Analysis of Recorded Sounds Relating to the Assassination of President John F. Kennedy. HSCA Volume VIII.

3 Ramsey NF, Alvarez LW, Chernoff H, Dicke RH, Elkind JI, Feggeler JC, Garwin RL, Horowitz P, Johnson A, Phinney RA, Rader C, and Sarles FW. Report of the Committee on Ballistic Acoustics. National Research Council (US). Washington: National Academy Press, 1982. Posted at http://www.nap.edu/catalog/10264.html.

4 Committee on Ballistic Acoustics, National Research Council (US). Reexamination of acoustic evidence in the Kennedy assassination. Science 1982; 218: 127–133.

5 Thomas DB. Echo correlation analysis and the acoustic evidence in the Kennedy assassination revisited. Science and Justice 2001; 41: 21–32.

6 O'Dell M. The acoustic evidence in the Kennedy assassination. 2003. Posted at http://mcadams.posc.mu.edu/odell/.

7 The MP3 files are posted at http://books.nap.edu/html/JFK_audio.

8 Agarwal RC, Garwin RL, and Lewis BL. Signal Processing Analysis of the Kennedy Assassination Tapes. Yorktown Heights, NY: IBM Research Report RC9771, 1982.

9 Oppenheim AV and Schafer RW. Discrete-Time Signal Processing. Englewood Cliffs, NJ: Prentice-Hall, 1989, p. 770.

10 Linsker, R. Separation of a Mixture of Acoustic Sources into its Components. 2001. U. S. Patent #6,317,703.

11 Bowles JC. The Kennedy Assassination Tapes: A Rebuttal to the Acoustical Evidence Theory. Transcript Channel 2. Posted at http://www.jfk-online.com/bowles8.html.