Supplementary Material

1. Corpus Analysis

We analyze the corpus of 100 games from four angles. First, we estimate the quality of BAPO. Second, we explore the mappings from broadcast timestamps into game timestamps and the resulting game segments in the broadcasts. Third, we look at the number of plays and possessors identified in the corpus. Finally, we quantify the average number of possessors per game.

Corpus Quality For the mapping of broadcast timestamps and game timestamps, we found 97% success rate in a sample of 500 mappings across the 100 games. For the accuracy of patterns to collect possession information, we manually checked 10 plays per game among the 100 games and found 100% success rate. And for the accuracy of transcripts, we manually checked 10 minutes per game for the 100 games and found 96% success rate, 3% word error rate due to commercials and background music, and 1% word error rate due to wrong transcriptions (noise, poor sound quality, etc.). We manually fixed the latter errors.

	Q1	Q2	Q3	Q4	All
% broadcast covering	19.63	27.00	21.99	28.09	96.71
% broadcast timestamp mapped to					
game	67.62	69.02	70.53	71.37	69.67
no_game	32.38	30.98	29.47	28.63	30.33
% game timestamps mapped to broadcast timestamp	94.86	94.95	95.55	94.89	95.05
# transitions (game, no_game) and (no_game, game)	65.98	78.04	61.88	73.44	279.40
Average game segment length					
broadcast seconds	32.03	38.59	38.35	42.54	37.54
game seconds	20.76	20.19	24.10	21.96	21.33
Average no_game segment length					
broadcast seconds	15.04	17.01	15.75	16.59	16.08

Table 1. Basic statistics for the corpus per quarter and all quarters. The average length of a broadcast is ≈2 hours. The broadcast segment covering a game quarter spans from the broadcast timestamp mapped to (minute 15, second 0) to the one mapped (minute 0, second 0), as game timestamps within a quarter decrease over (real) time. The maximum number of game timestamps per quarter that can be mapped is 900 (15 minutes, 60 seconds per minute).

#players with ball	Quarter 1		Quarter 2		Quarter 3		Quarter 4		All Quarters	
	#plays	%poss	#plays	%poss	#plays	%poss	#plays	%poss	#plays	%poss
0	76	n/a	54	n/a	41	n/a	46	n/a	217	n/a
1	1,943	48.9	2,337	51.4	1,934	49.3	2,321	52.7	8,535	50.7
2	1,323	67.5	1,677	68.2	1,409	67.2	1,551	67.9	5,960	67.7
>2	83	75.4	108	76.5	97	74.5	132	70.9	420	74.5
Any	3,425	55.7	4,176	58.1	3,481	56.8	4,050	58.5	15,132	57.4

Table 2. Number of play descriptions depending on the number of players who had the ball, and percentage of players who had the ball out of all players mentioned in the description.

Broadcast and Game Timestamps We show a numeric description of the mapping between broadcast and game timestamps, and the resulting game segments in Table 1. The majority of the ≈2-hour broadcast (96.71%) covers the period between the

initial and ending game timestamps ((quarter: 1, minute: 15, second: 0) and (quarter: 4, minute: 0, second: 0)). Regarding quarters, the broadcast segments covering the second and fourth quarter (before halftime and before the end of the game) are longer than the rest (27–28% vs. 20–22%). Unsurprisingly, we find that only 70% of the broadcast is broadcasting the game live. The rest are commercials, time outs and other interruptions.

The methodology described to map video timestamps into game timestamps finds 95% of all possible mappings in all quarters. This number may seem low, but manual inspection of 200 unmapped game timestamps across the 100 games reveals that 98% of them are skipped in the broadcast game and no_game and thus are not errors. There is a substantial amount of transitions between game and no_game broadcast segments, especially in the second and fourth quarters (78 and 73), showing that game interruptions in these quarters are not just longer—there are actually more interruptions.

Game segments are longer than no_game segments in broadcast time across all quarters (average: 38 vs. 16 seconds). Broadcasts take longer to cover game segments as the game progresses, in particular in the second and fourth quarters. Indeed, broadcasts take double the game clock to cover a game segment in these two quarters.

Plays and Possessors Our corpus includes 15,132 plays from the 100 games (Table 2). The number of plays stays relatively uniform across all quarters, although we again observe a larger number in the second and fourth quarter. The patterns used to extract which player has the ball are successful (i.e., at least one pattern matches) in the majority of plays (all but 217 plays, 98.6%). We examined the descriptions of these 217 plays manually and discovered that there is no possession information to be extracted. For example, no possession information can be extracted from the description *Timeout # 1 by Arizona Cardinals*.

We identify one possessor in most plays (8,535, 56%), but we identify two or more possessors in a substantial amount of plays (5,960 and 420 plays, 39.4% and 2.8%). Finally, we note that play descriptions often mention players that do not have the ball, as exemplified in Table 1. We observe that the chance of a player who is mentioned to be in possession of the ball increases with the number of possessors identified. On average, there is one additional player mentioned in addition to the possessor: 2 players when there is one possessor (50.7% chance), and 3 players when there are two possessors (67.7% chance).

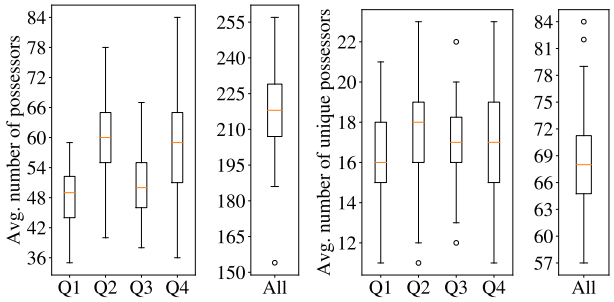


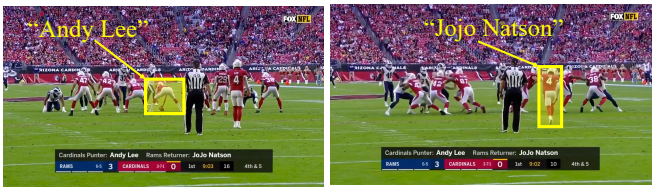
Figure 1. Distribution of possessors (left: all possessors, right: unique possessors) in the corpus of 100 football games (per quarter and all four quarters).

Possessors per game. Figure 1 shows the distribution of possessors per game (all possessors and unique possessors). Each boxplot shows the minimum, 25th percentile, median (or 50th percentile), 75th percentile and maximum of the number of possessors across the 100 games. Recall that we identify more than one possessor in many plays (42.2%, Table 2) thus the median number of possessors is higher than the median number of plays per game.

The number of possessors identified per game ranges from 186 to 257 except for an outlier with 154 possessors. The median is substantial (218 possessors), and the statistics per quarter show that the corpus identifies more possessors in the second and fourth quarter, as the number of plays per quarter would suggest (Table 2). Regarding unique possessors, we observe that many players are in possession of the ball many times during a game: the number of unique possessors ranges from 57 to 84, and the median is 68. There is, however, a substantial amount of players that have the ball at some point during a football game; it is not the case that only a handful of players have the ball in a game. Additionally, the number of possessors is substantially larger than the amount of scoring events. On average, we observe only $\approx$10 scoring events, fouls or other important game events targeted in previous work (Section 2).

2. Case study

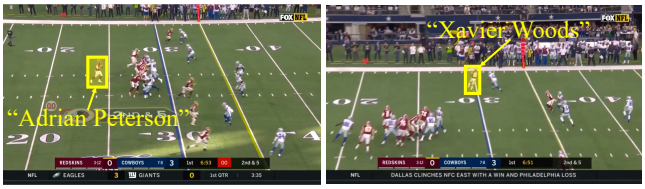
To further investigate how the three modalities can help solve the football possession task, we randomly select two examples and show the comparisons between ground truth and the predicted results of BAPOTer.



Transcript: "…so here comes Andy Lee, the veteran Potter who has pulled off of fake is successful when I get to the Buccaneers, Jojo Natson back to receive for the Rams data gets out of 25."

Ground truth: Andy Lee, Jojo Natson
BAPOTer: Andy Lee, Jojo Natson
Ground truth: Arizona Cardinals, Los Angeles Rams
BAPOTer: Arizona Cardinals, Los Angeles Rams

(a) Correct Example



Transcript: "… Cowboys get another check away, second fumble of the Year lost by Adrian Peterson as Malcolm Smith in his second week with Dallas forced it and Xavier Woods comes away."

Ground truth: Case Keenum, Adrian Peterson, Xavier Woods
BAPOTer: Adrian Peterson, Xavier Woods
Ground truth: Washington Redskins, Dallas Cowboys
BAPOTer: Washington Redskins, Dallas Cowboys

(b) Partially Correct Example

Figure 2. Randomly selected examples from BAPO. The transcripts and corresponding video frames, ground truth and prediction results are provided. (a) is a correct prediction while (b) is a partially correct prediction. We underline the mentioned possessors in the transcript and mark them with bounding box in the video frames.

In the first example (Figure 2a), given the video, audio and text, BAPOTer correctly predicts the two players and two teams who possess the ball. However, in the second example (Figure 2b), only two possessors are correctly predicted while there are three possessors in the ground truth. The reason might be that only Adrian Peterson and Xavier Woods are mentioned in the audio and transcript. This is consistent with our experimental results that demonstrated that language is the most important modality. This example tells us that there are possessors appearing in the video but not mentioned in the audio or text. Thus, the video is sometimes a more complete depiction than the audio or transcript.

3. Details of Regular Expression Patterns

Pattern	%	Example
Player$_1$ pass complete short left / middle / right to Player$_2$ for [. . .]	23.7	Kyler Murray pass complete short right to Larry Fitzgerald for 9 yards (tackle by *Jessie Bates* and *Geno Atkins*)
Player$_1$ pass incomplete short [. . .]	9.0	Carson Wentz pass incomplete short right intended for *Boston Scott*
Player$_1$ up the middle for [. . .]	8.2	Adrian Peterson up the middle for 5 yards (tackle by *Robert Quinn*)
Player$_1$ left / right end for [. . .]	8.0	David Johnson left end for 1 yard (tackle by *Geno Atkins* and [. . .]
Player$_1$ left / right tackle for [. . .]	6.8	Miles Sanders right tackle for -1 yards (tackle by *Julian Love*)
Player$_1$ left / right guard for [. . .]	6.4	Saquon Barkley left guard for -3 yards (tackle by *Anthony Rush*)
Player$_1$ kicks off [0-9]+ [. . .]	6.1	Zane Gonzalez kicks off 65 yards, touchback
Player$_1$ pass incomplete deep left / right intended for [. . .]	5.9	Daniel Jones pass incomplete deep right intended for *Sterling Shepard* (defended by *Avonte Maddox*)
Player$_1$ pass complete deep left / right to Player$_2$ for	5.0	Daniel Jones pass complete deep right to Darius Slayton for 33 yards (tackle by *Rasul Douglas*)
Player$_1$ kicks extra point good	4.8	Jake Elliott kicks extra point good
Player$_1$ punts [0-9]+ [. . .]	3.7	Cameron Johnston punts 43 yards
Player$_1$ sacked by PLAYER2	3.5	Daniel Jones sacked by *Timmy Jernigan* for -10 yards
[...] returned by Player$_1$ for [0-9]+	2.6	Mitch Wishnowsky kicks off 66 yards, returned by Reggie Bonnafon for 24 yards
Player$_1$ [...] yard field goal good	2.1	Jake Elliott 31 yard field goal good
Player$_1$ kicks onside [...]	1.8	Brandon McManus kicks onside 9 yards,
[...] is intercepted by PLAYER$_1$	1.0	Josh Allen pass deep middle is intercepted by Xavien Howard at MIA-0 and returned for no gain
[...] muffed catch by PLAYER$_1$	0.7	Matt Wile punts 37 yards, muffed catch by Danny Amendola
[...] recovered by PLAYER$_1$	0.5	*Jameis Winston* fumbles, recovered by Mike Evans at NYG–3, touchdown
[...] fair catch by PLAYER$_1$	0.2	Sam Koch punts 39 yards, fair catch by Desmond King at LAC-12

Table 3. The full table of frequent patterns used to extract which players had the ball from full play-by-play textual descriptions. Underlining indicates the player or players who had the ball, and italics indicates other players mentioned.