

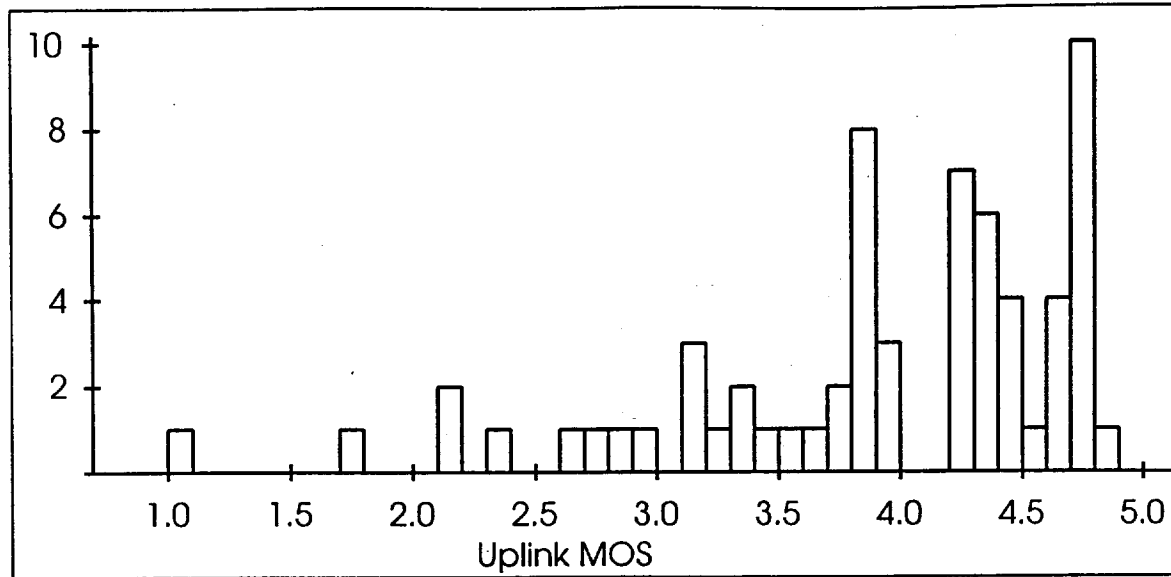


Figure 7.3.1.3 Frequency Distribution of Uplink MOS

Figure 7.3.1.4 compares all MOS scores against one objective measure: average RSS (dBm). Figures 7.3.1.5 and 7.3.1.6 compare MOSs with average WER on the downlink and uplink, respectively.

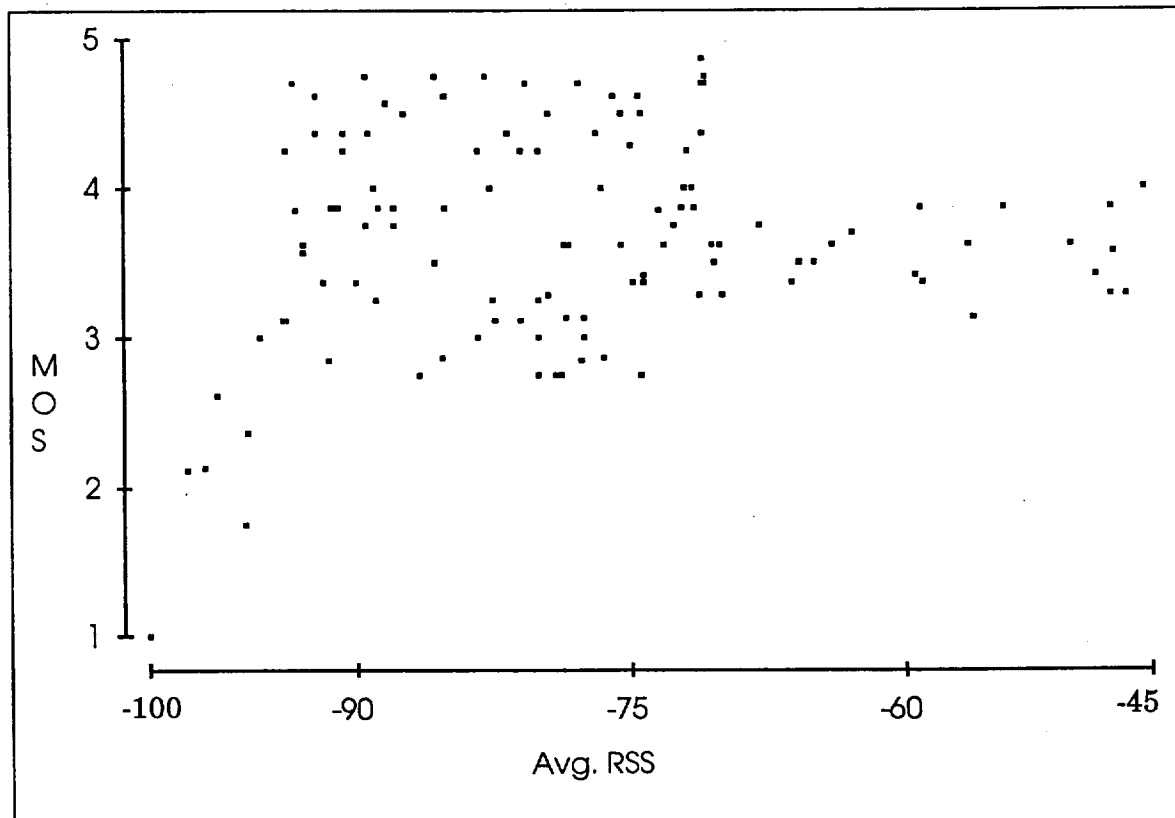


Figure 7.3.1.4

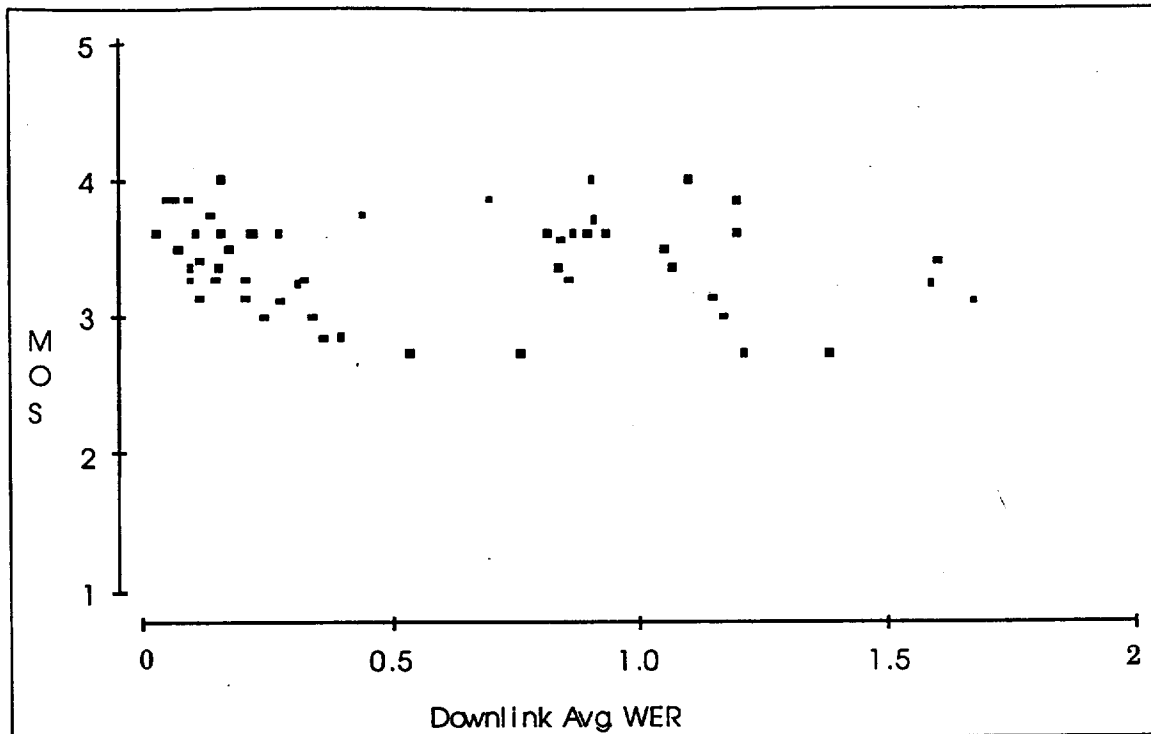


Figure 7.3.1.5

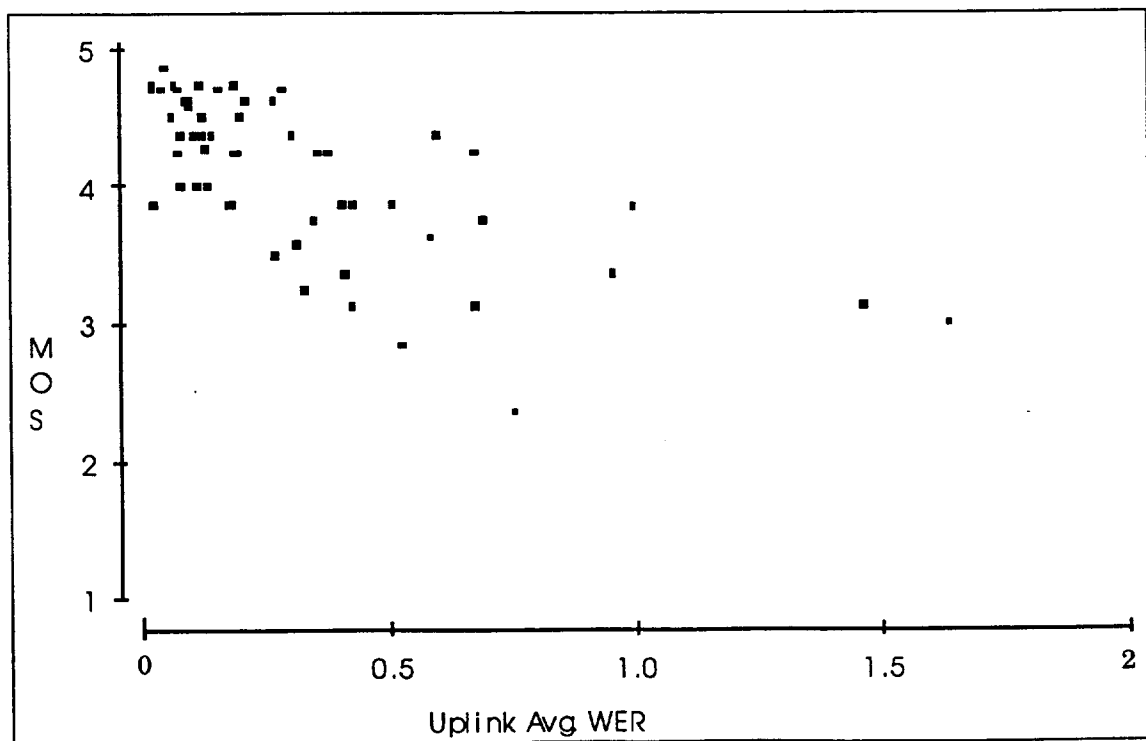


Figure 7.3.1.6

Figures 7.3.1.5 and 7.3.1.6 indicate that while the uplink samples show a relatively high correlation between high MOS and low WER, the downlink samples do not.

Further statistical analysis on MOS scores and engineering data from the downlink and uplink runs illustrates the differences between the two links.

Figure 7.3.1.7 is a Pearson Product-Moment correlation table of downlink data and outlines the correlation coefficients of MOS vs. RSS and WER. The results reveal that there is a linear correlation between MOS and EL and between MOS and RSS, but surprisingly little correlation between MOS and WER or WER and RSS.

Pearson Product-Moment Correlation				
	MOS	EL	Avg. RSS	Avg. WER
MOS	1.000			
EL	0.526	1.000		
Avg. RSS	0.475	0.140	1.000	
Avg. WER	-0.052	0.079	-0.139	1.000

Figure 7.3.1.7 Downlink

A Spearman Rank Correlation was performed to analyze the ranks of MOS vs. the ranks of the various engineering measures, in order to consider possible non-linear relationships. Figure 7.3.1.8 diagrams the results.

Spearman Rank Correlation				
	MOS	EL	Avg. RSS	Avg. WER
MOS	1.000			
EL	0.540	1.000		
Avg. RSS	0.549	0.239	1.000	
Avg. WER	-0.246	-0.281	-0.423	1.000

Figure 7.3.1.8 Downlink

The Spearman Rank Correlation does illustrate that the non-linear correlations are stronger than the linear correlations, but they are still not as high as expected.

Figure 7.3.1.9 is a Pearson Product-Moment correlation table of uplink data and outlines the correlation coefficients of MOS vs. RSS and WER. The results reveal some correlations between averaged engineering measures and MOS, the highest being WER at -0.550 (the negative sign denotes higher MOS corresponds to lower WER, as expected). Therefore, there does appear to be a linear relationship between MOS and the engineering measures. Notice that the correlation between MOS and WER is stronger than the correlation between WER and RSS.

Pearson Product-Moment Correlation				
	MOS	EL	Avg. RSS	Avg. WER
MOS	1.000			
EL	0.849	1.000		
Avg. RSS	0.627	0.564	1.000	
Avg. WER	-0.550	-0.571	-0.397	1.000

Figure 7.3.1.9 Uplink

There is a linear relationship between MOS and the engineering measures, but there may be stronger relationships that are non-linear, as a Spearman Rank Correlation table might reveal. A Spearman Rank Correlation was performed to analyze the ranks of MOS vs. the ranks of the various engineering measures, in order to consider possible non-linear relationships. Figure 7.3.1.10 diagrams the results. There appear to be strong non-linear relationships between RSS, WER and MOS. Note that the non-linear relationship between MOS and WER is stronger than the linear relationship between the two.

Spearman Rank Correlation				
	MOS	EL	Avg. RSS	Avg. WER
MOS	1.000			
EL	0.735	1.000		
Avg. RSS	0.647	0.642	1.000	
Avg. WER	-0.783	-0.669	-0.737	1.000

Figure 7.3.1.10 Uplink

Note, the engineering values were averaged over the entire voice sample length. By analyzing the engineering data within the voice sample time frame, further insight may be gained on the behavior of MOS. By considering the minimum and maximum values for WER and RSS over the length of a given sample, we might see higher correlations to MOSs.

7.3.2 Expert Listener Data Analysis and Discussion

Recall, the expert listener ratings were performed on a 3 point scale, and were made in an attempt to emulate listener panels' ratings of acceptability (vs. MOS).

Figures 7.3.2.1 and 7.3.2.2 show the relationship between expert listener (EL) ratings and listener responses to question three, or %acceptability [p(A)]. Percent acceptability is the percent of listeners rating a given sample as acceptable.

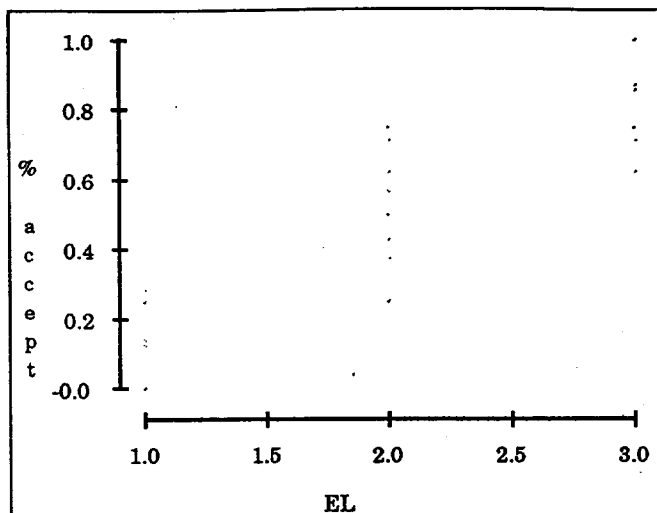


Figure 7.3.2.1

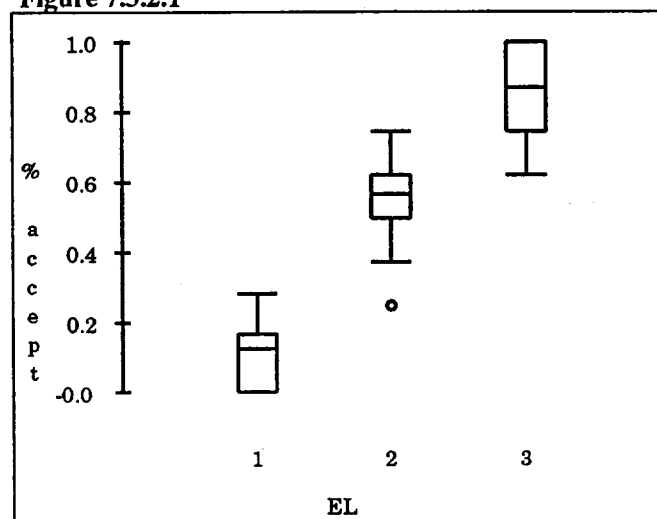


Figure 7.3.2.2

Figure 7.3.2.1 shows the %acceptability, or $p(A)$, data in relation to EL categories. Figure 7.3.2.2 summarizes the same data. The boxes represent the middle half of the data (from the 25th percentile to the 75th percentile). The line intersecting the boxes are the median $p(A)$ s for each of the EL ratings. The lines extending out of the boxes depict the spread of the data. Our goal was for Expert Listener scores to accurately predict Listener Panel % acceptability scores, and this goal was met for 95% of the 120 voice samples.

When we examine the relationship between listener responses to question 1 (MOS) and question 3 [$p(A)$] we see that they are strongly correlated as we would expect. The correlation coefficient between MOS and $p(A)$ is 0.871. The EL ratings are strongly correlated with MOS, with a coefficient of 0.749, and strongly correlated with percent acceptability, or $p(A)$, with a correlation coefficient of 0.9 (see Figure 7.3.2.3).

Pearson Product-Moment Correlation			
	MOS	% accept	EL
All MOS	1.000		
All p (A)	0.871	1.000	
All EL	0.749	0.900	1.000

Figure 7.3.2.3

It is possible that EL ratings, average RSS, and average WER can be predictors of MOS when all are taken together. A multiple regression analysis was completed in order to determine if MOS is related to a combination of all the above factors. Interactions among the various factors were not considered in the analysis.

Since previous statistical analysis revealed statistically significant differences between uplink and downlink data, separate regression analyses were performed on each set of data. The regression analysis of the downlink data is as follows:

Dependent variable is: MOS				
R squared = 44.1% R squared (adjusted) = 40.9%				
s = 0.2677 with 56 - 4 = 52 degrees of freedom				
Source	Sum of Squares	df	Mean Square	F-ratio
Regression	2.94508	3	0.981694	13.7
Residual	3.72719	52	0.071677	
Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	3.51377	0.3152	11.1	§ 0.0001
EL	0.328797	0.0734	4.48	§ 0.0001
Avg. RSS	0.012801	0.0034	3.82	0.0004
Avg. BER	-5.59928e-3	0.0181	-0.310	0.7577

Figure 7.3.2.4 Downlink Regression Analysis Results

The result is an R squared (adjusted) of 40.9%. In other words, 40.9% of the variance in MOSs can be explained by EL, RSS, and WER. The F-ratio is rather low (13.7), revealing all factors may not reliably account for the 40.9% of variance. By including all of the factors into the equation, we see some predictive power over MOS, but it is less than anticipated. (see Figure 7.3.2.4 above).

A second regression analysis was performed on the downlink data, this time with only the Expert Listener (EL) variable. The result was an R squared (adjusted) of 26.3%. This demonstrates that a relatively large amount of the variance in MOSs can be explained by EL alone, and the RSS and WER variables are much less predictive.

Dependent variable is: MOS				
R squared = 27.6% R squared (adjusted) = 26.3%				
s = 0.2990 with 56 - 2 = 54 degrees of freedom				
Source	Sum of Squares	df	Mean Square	F-ratio
Regression	1.84318	1	1.84318	20.6
Residual	4.82909	54	0.089428	
Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	2.52298	0.2001	12.6	Š 0.0001
EL	0.366604	0.0808	4.54	Š 0.0001

Figure 7.3.2.5 Downlink Regression Analysis Results (only EL variable)

The regression analysis of the uplink data is as follows:

Dependent variable is: MOS

R squared = 75.7% R squared (adjusted) =
s = 0.4213 with 64 - 4 = 60 degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	33.0887	3	11.0296	62.1
Residual	10.6483	60	0.177471	
Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	3.57471	0.7903	4.52	Š 0.0001
EL	0.808950	0.1019	7.94	Š 0.0001
Avg. RSS	0.019951	0.0074	2.69	0.0092
Avg. BER	-0.036030	0.0384	-0.939	0.3514

Figure 7.3.2.6 Uplink Regression Analysis Results

The result is an R squared (adjusted) of 74.4%. In other words, 74.4% of the variance in MOSs can be explained by EL, RSS, and WER. The F-ratio is relatively high (62.1), revealing all factors should reliably account for the 74.4% of variance. By including all of the factors into the equation, we see some predictive power over MOS. (see Figure 7.3.2.4 above).

A second regression analysis was performed on the uplink data, this time with only the Expert Listener (EL) variable. The result was an R squared (adjusted) of 71.6%. This demonstrates that a great deal of the variance in MOSs can be explained by EL alone, and the RSS and WER variables are much less predictive.

Dependent variable is: MOS				
R squared = 72.1% R squared (adjusted) = 71.6%				
s = 0.4437 with 64 - 2 = 62 degrees of freedom				
Source	Sum of Squares	df	Mean Square	F-ratio
Regression	31.5284	1	31.5284	160
Residual	12.2086	62	0.196913	
Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	1.36021	0.2093	6.50	§ 0.0001
EL	0.996504	0.0788	12.7	§ 0.0001

Figure 7.3.2.7 Downlink Regression Analysis Results (only EL variable)

7.3.3 Other Listener Comments

By gathering listeners' comments from post-test questionnaires, more information is available about the nature of MOSs. Namely, it is evident from questionnaires that there are several types of distortions in quality possible in the TAG 3 voice samples according to listeners:

- 1) "static" (5)
- 2) "background noise" (5)
- 3) "popping noises" (2)

Note that the words in quotes are the exact words used by listeners, and the number in parentheses indicates the number of listeners who made that comment.

The nature of these distortions are likely judged differently by different listeners. Intelligibility and speaker recognition are two main aspects of perceived quality. In the case of most of the voice samples, intelligibility remained high. As a result, overall MOSs were high.

7.3.4 Handover Voice Quality Results

Two separate sets of handover tests were conducted: one at pedestrian walking speed and one at various automobile speeds.

The pedestrian tests consisted of 14 runs which were conducted within the same grid which was used for the Quasi-stationary points. Expert listener ratings were given by Jaye Matthews at four second intervals over the duration of each run. Voice quality remained acceptable over the length of each uplink run, with no degradation until the call was dropped. None of the uplink handovers were distinguishable to the expert listener. Voice quality was primarily acceptable over the length of each downlink run, however 8 of the 14 runs did have brief (< 3 seconds) periods of marginally acceptable voice quality. These periods were not associated with handover. None of the downlink handovers were distinguishable to the expert listener.

The high speed (automobile) test was conducted in a van traveling at the following rates of speed: 20 mph, 30 mph, 40 mph, 50 mph, 60 mph. Two runs were conducted at each rate of speed, for a total of 10 runs. Expert listener ratings were given by Jaye Matthews at four second intervals over the duration of each run. On the uplink samples, voice quality remained acceptable over the length of each run, with no degradation until the call was dropped. Approximately 28% of uplink handovers were distinguishable to the expert

listener as a 200 - 400 millisecond period of muting. The remaining uplink handovers were not at all distinguishable to the expert listener. On the downlink samples, 40% of the runs had voice quality degradation as the call approached the edge of coverage. Approximately 50% of downlink handovers were distinguishable to the expert listener as a 200 - 400 millisecond period of muting. The remaining downlink handovers were not at all distinguishable to the expert listener. Maps 7.3.6.1 through 7.3.6.20 illustrate the high speed handover voice quality for each run, uplink and downlink.

7.3.5 Voice Quality Conclusions

Overall the voice samples were rated favorably, with 88% of the samples being rated between fair (3.0) and excellent (5.0). The average of all MOSs is 3.68, with a standard deviation of 0.6975. On the whole, average WERs were between 0.01% and 4%, with three outliers at 6.2, 10.9 and 15.3%, while MOSs varied from 1.0 to 4.875.

There was a statistically significant difference in MOSs for uplink and downlink when paired MOSs were analyzed. Average uplink MOS was 3.914 and average downlink MOS was 3.413. The median uplink MOS was 4.25 and the median downlink MOS was 3.43.

Voice quality generally remained acceptable during pedestrian and high speed handover runs. None of the pedestrian speed handovers were distinguishable to the expert listener. Twenty-eight percent of uplink high speed handovers and 50% of downlink high speed handovers were distinguishable to the expert listener.

By gathering listeners' comments from post-test questionnaires, more information is available about the nature of MOSs. Namely, it is evident from questionnaires that there are several types of distortions in quality possible in the TAG 3 voice samples according to listeners:

- 1) "static" (5)
- 2) "background noise" (5)
- 3) "popping noises" (2)

Appendix 7.A: Voice Quality Instructions

- Good afternoon/evening and thank you for coming. My name is Jaye Matthews, your test administrator.
- Tonight/today you will be listening to a series of recordings and telling us what you think about them.
- Before we begin, I would like to point out some features of our lab. There are two video cameras set up in the room. I will be recording this session. The resulting tapes would be used for research purposes only, and will not be distributed for advertising, etc. Is it okay with everyone if I continue recording? Also in our lab we have a two-way mirror. There may be other product designers and testers observing this session, to see how things are progressing and how the product is performing.
- As I already mentioned, you will be listening to a series of recordings and offering your opinions on them. For each recording we would like you to compare the sound quality of the recording to the service you have on your home phone and then enter your rating of that sound into the computer in front of you. For each, you will rate 3 things:
 - 1) What do you think of the **QUALITY** of the sound? That is, how good is the sound of the voice being captured by the recording? Remember, we'd like you to compare the samples to what you might hear at home.
 - 2) How annoying are any additional sounds you might hear? That is, How much do any additional sounds interfere with your ability to understand what is being said?
 - 3) Would you find this recording **ACCEPTABLE** as portable phone service? What we mean by "portable" is that you can take this phone with you as you travel around town and make calls from places like shopping centers, parking lots, and sporting events. Any questions so far?
- In front of you is a telephone handset which is connected to my computer here. I would like you to hold the handset up to your ear and listen to the sound I will play.
- I will announce the number of the sound you are about to hear and that num**WER** should match the number at the top of the screen. You should all have a #1 at the tops of your screens.
- You can use the mouse or the keypad to enter your ratings. Once you have entered your rating for one question the computer will move on to the next question for you. If you need to change an answer, you can tab through to go back to the answer you wish to change.
- The first few will be for practice, and you will have a chance to practice answering the questions with the computer. I will play the first practice sample sound for you now. **SOUND**.
- **NOW**, for practice, I would like you to rate that sound as Excellent for the overall quality of the sound. You can do this by entering a 1 on your keyboard or by pointing to the circle next to the word Excellent using the mouse and clicking the button on the mouse.
- Was everyone able to do that?
- Now please do the same to say that the additional sounds were Not at all Annoying and that the sound was Acceptable.
- Was everyone able to do that?
- Once you entered an answer for each question, A picture appeared in the bottom of the window which says "DONE". You can indicate that you are finished by clicking on "Done" with the mouse OR by hitting the return key on your keyboard.
- You can use either the mouse or the keyboard, whichever you like best.
- You will be hearing the same sentences in each sound bite. I realize that they are a bit unusual and at the end I will explain why we use these sentences. In some of the samples you will hear the male voice first, and in some the female voice first.

- Now try the next practice sample. SOUND
- Was everyone able to do that? Please rate that one as Bad by selecting #7. Any questions?
- If you have any questions as we are working, please feel free to speak-up. Also, If you are having difficulty using the computer please tell me.
- About half way through the voice segments we will take a 15 minute break and then finish up the rest of the samples.
- I would like you to rate the rest of the samples on you own.

Appendix 7.B: Computer Response Questions

How would you rate the overall QUALITY of the sound?

1. Excellent
2. Good
3. Fair
4. Poor
7. Bad

How annoying are any ADDITIONAL sounds you might hear?

1. Not at all annoying
2. Slightly annoying
3. Annoying
4. Very annoying
7. Extremely Annoying

Would this be ACCEPTABLE as portable service?

1. Acceptable
2. Not acceptable

Appendix 7.C: Post-Test Questionnaire

1) What is your gender?

☐ MALE
☐ FEMALE

2) What is your age?

☐ Under 20	☐ 35-44
☐ 21-24	☐ 45-54
☐ 25-34	☐ 55 or above

3) What is your occupation? _____

4) How much computer experience do you have?

☐ None	☐ Use once/week
☐ Used a few times before	☐ Use every day
☐ Use once/month	

Did you ever work with a computer on a regular basis in a previous job? If so,
how much? _____

5) Have you ever used a cordless phone (works on standard phone jack)?

☐ YES ☐ NO

Do you OWN a cordless phone?

☐ YES and use it more than 50% of the time
☐ YES and use it less than 50% of the time
☐ NO

IF YES. How does the sound quality of the samples you heard here compare to the sound on a cordless
phone?

IF YES. Who made the phone? _____

6) Have you ever used a cellular phone? ☐ YES ☐ NO

If so, how often:

☐ everyday
☐ once or twice per week
☐ once or twice per month
☐ once or twice per year
☐ only once or twice before

Have you ever received a call from someone using a cellular phone?
(that you know of) ☐ YES ☐ NO

If so, how often:

- ☐ everyday
- ☐ once or twice per week
- ☐ once or twice per month
- ☐ once or twice per year
- ☐ only once or twice before

If you have used a cellular phone or received a call from a person using a cellular phone, how does the sound quality of the samples you heard tonight compare to the sound of a cellular phone?

7) Would your ratings of the Acceptability of these sounds be different if we told you that this would NOT be a portable phone?

☐ YES ☐ NO

IF YES, How would they differ and why?

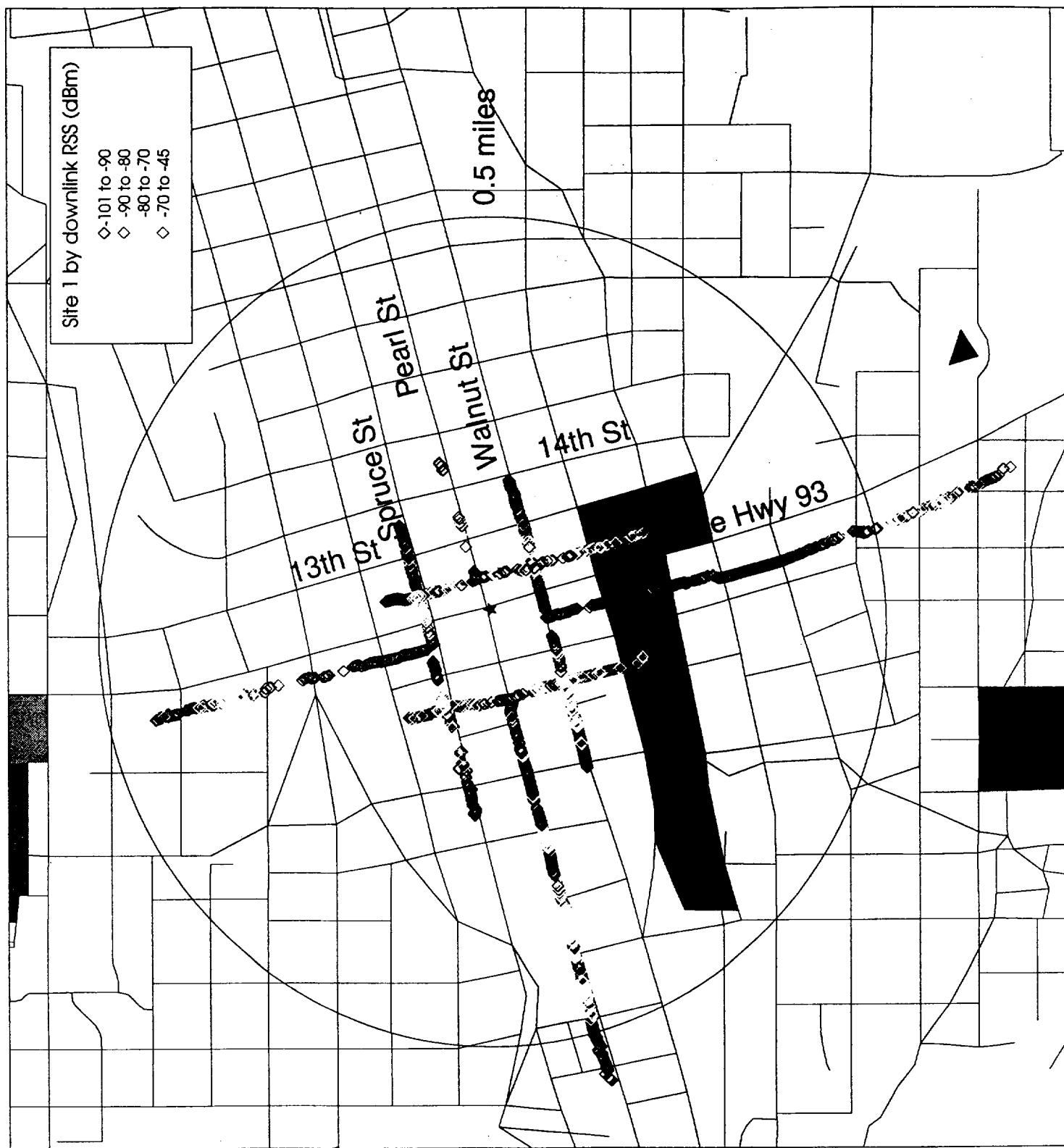
8) How does the sound from these recordings differ from the sound on your home phone?

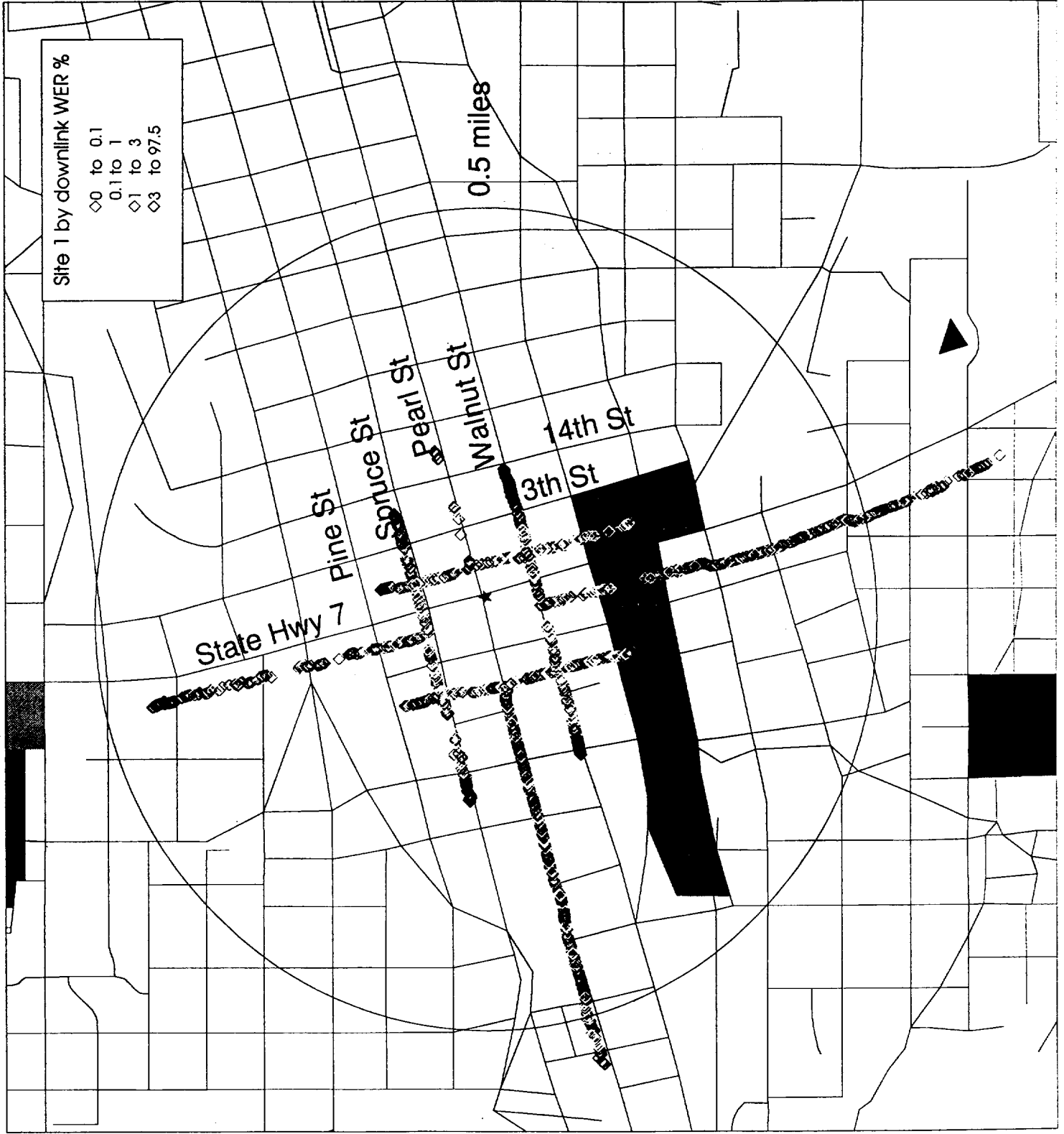
9) Any other comments?

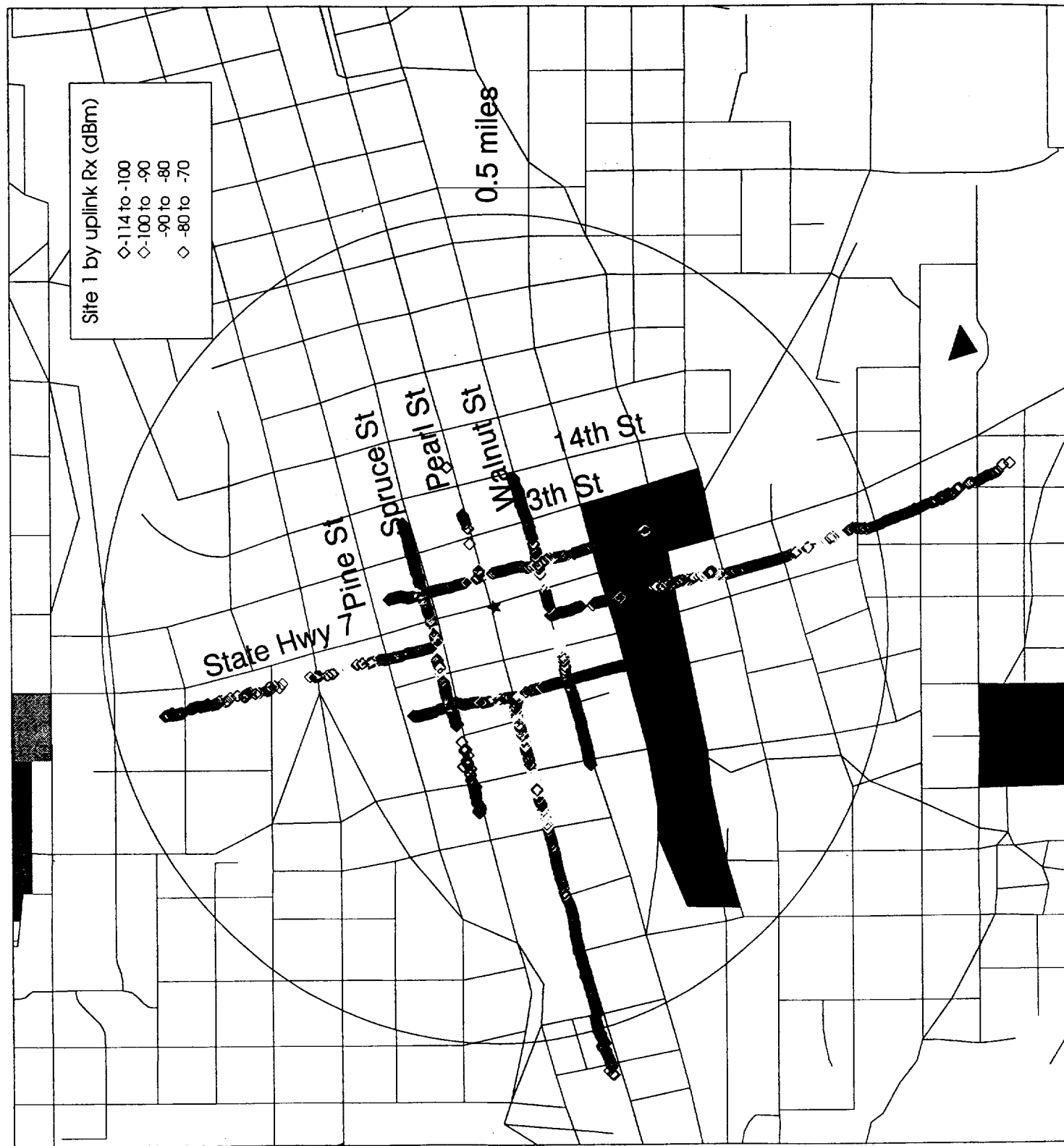
Appendix A. Coverage and Voice Quality Maps

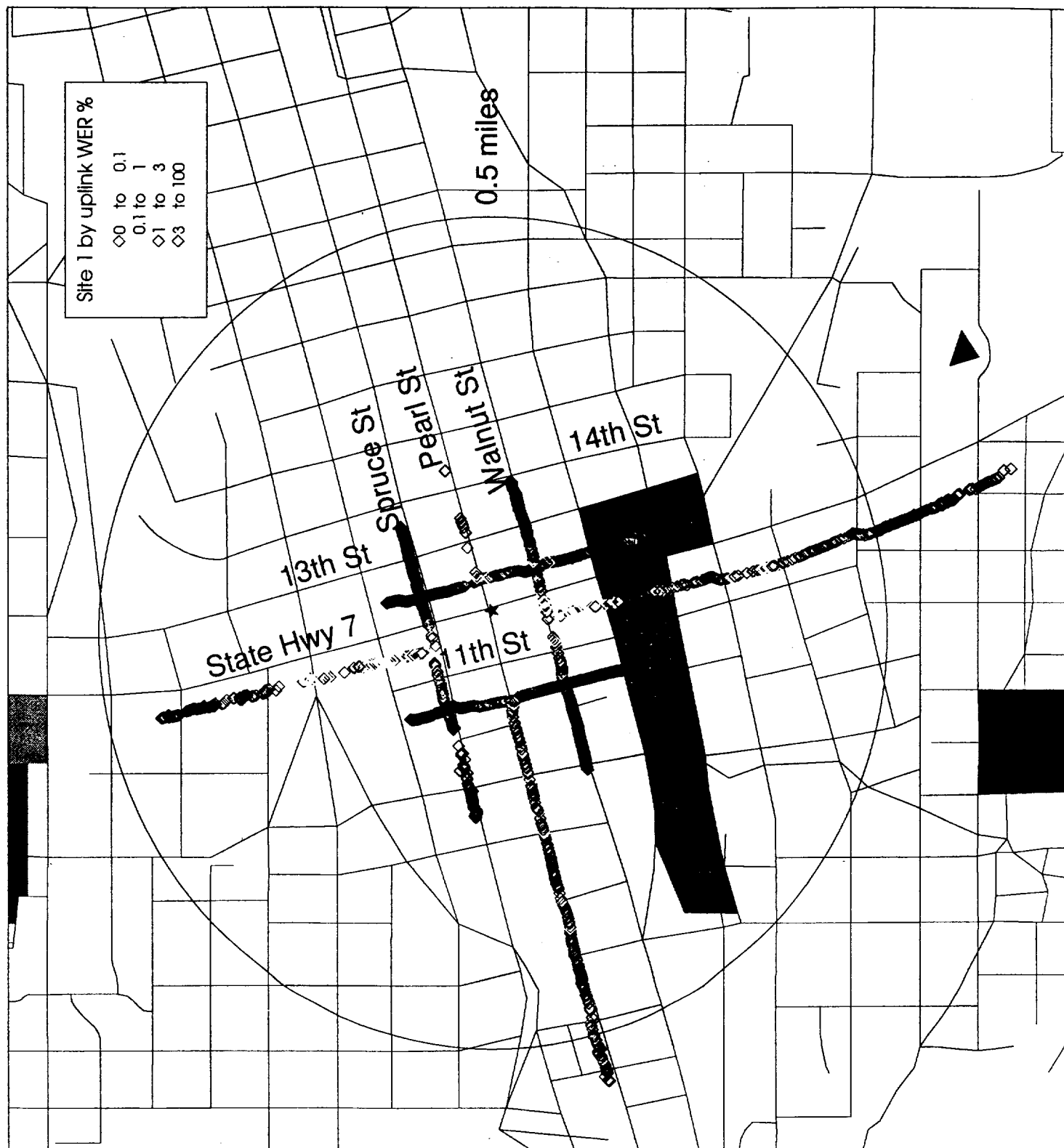


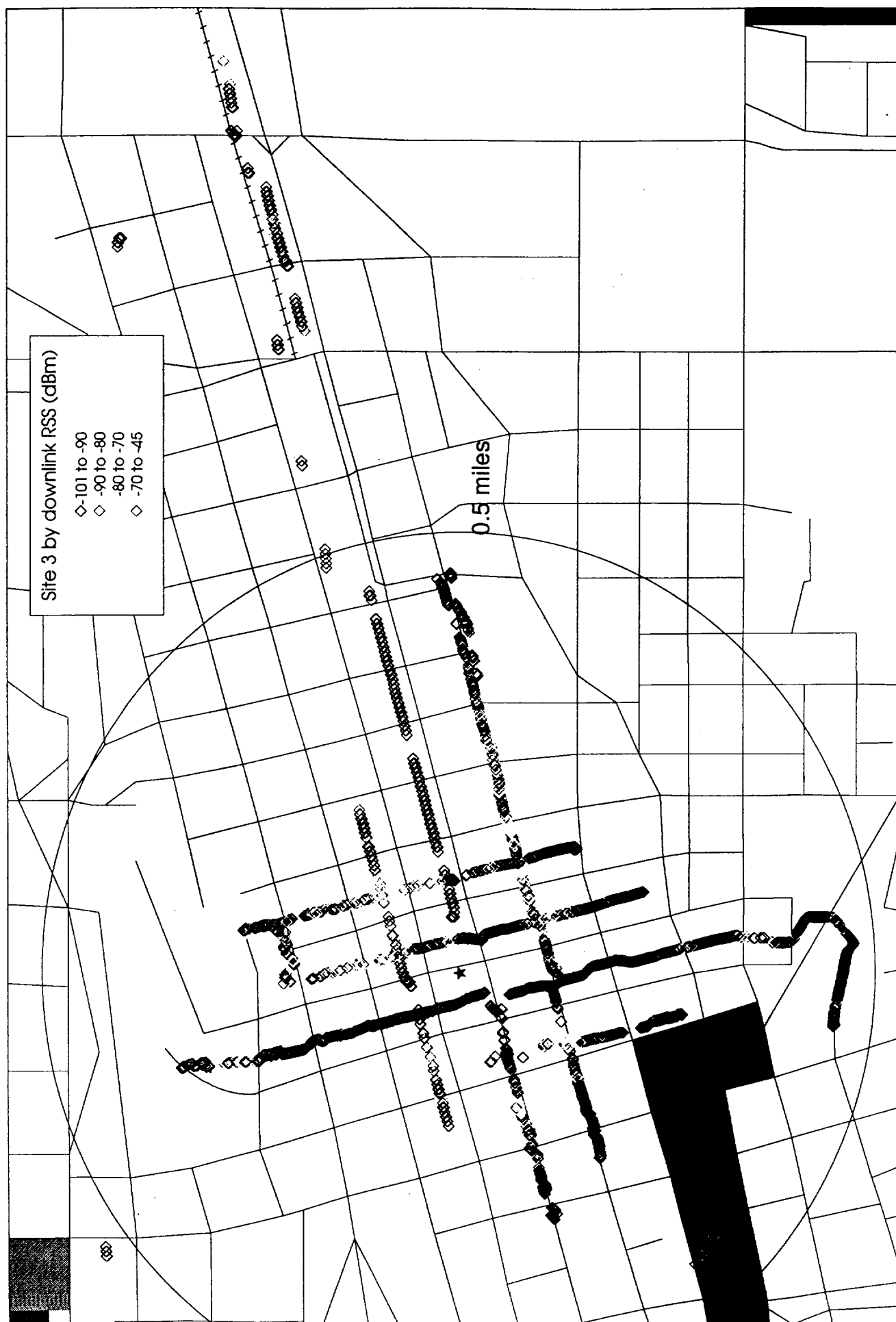
Map 4.2 USGS Map of Boulder



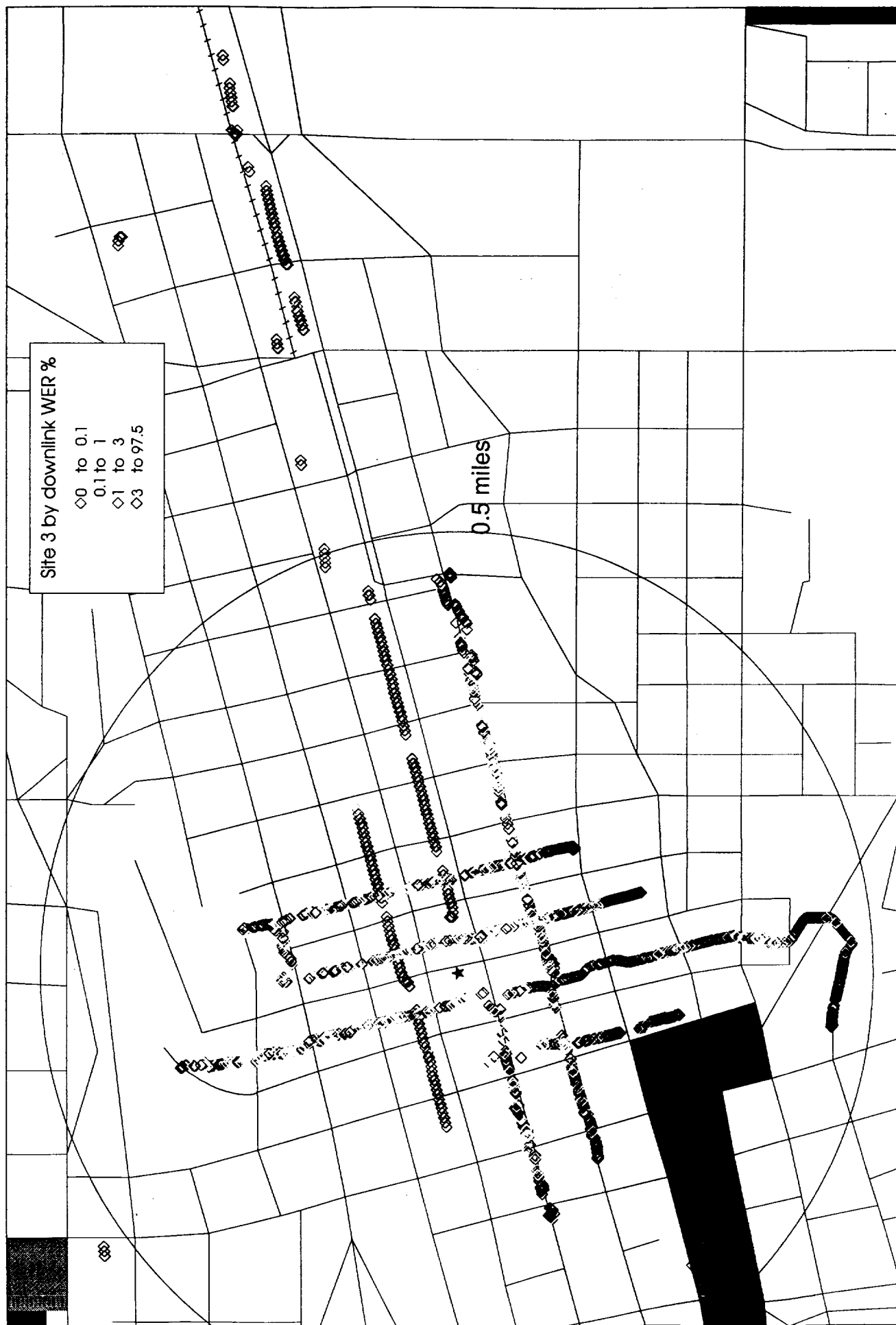




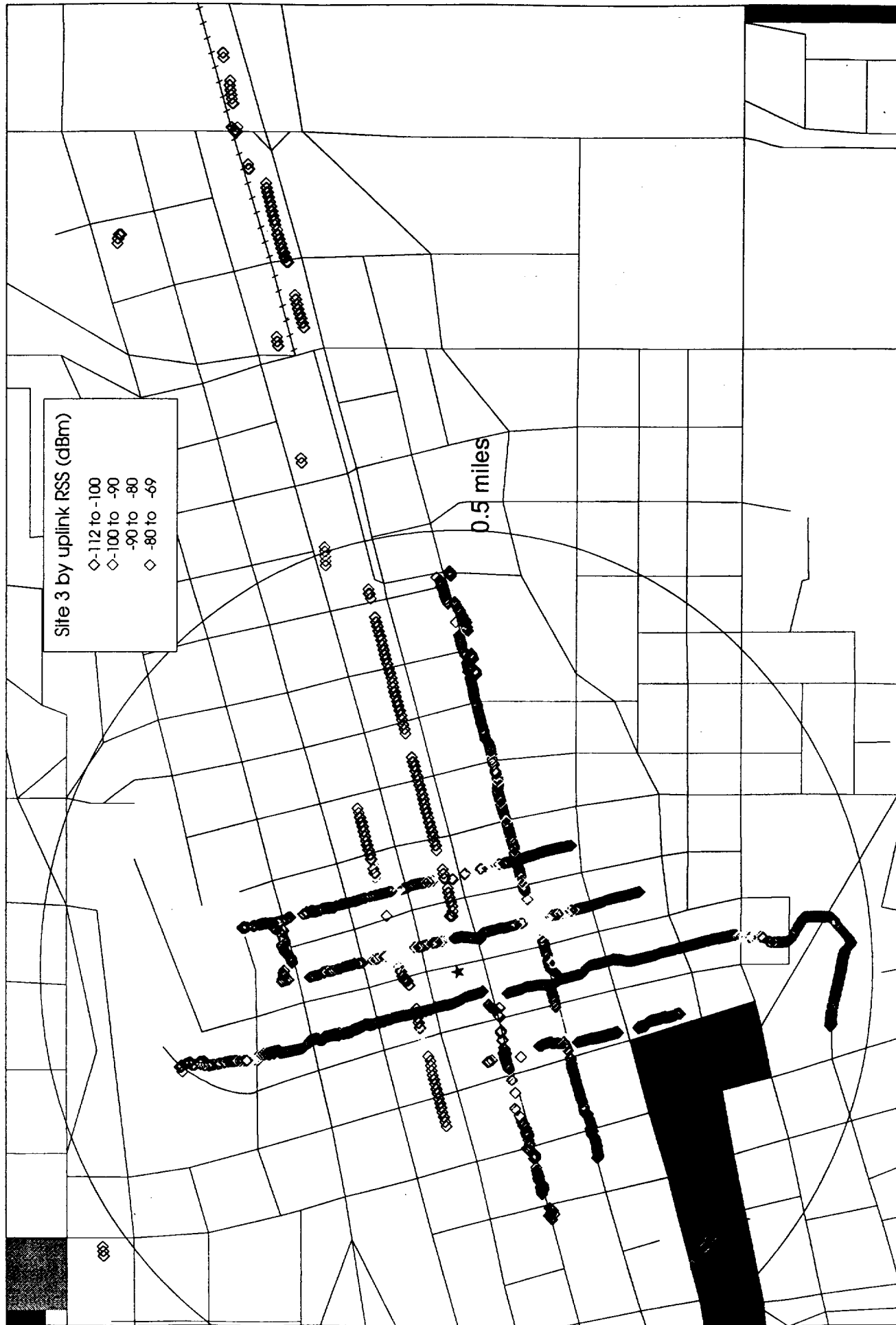




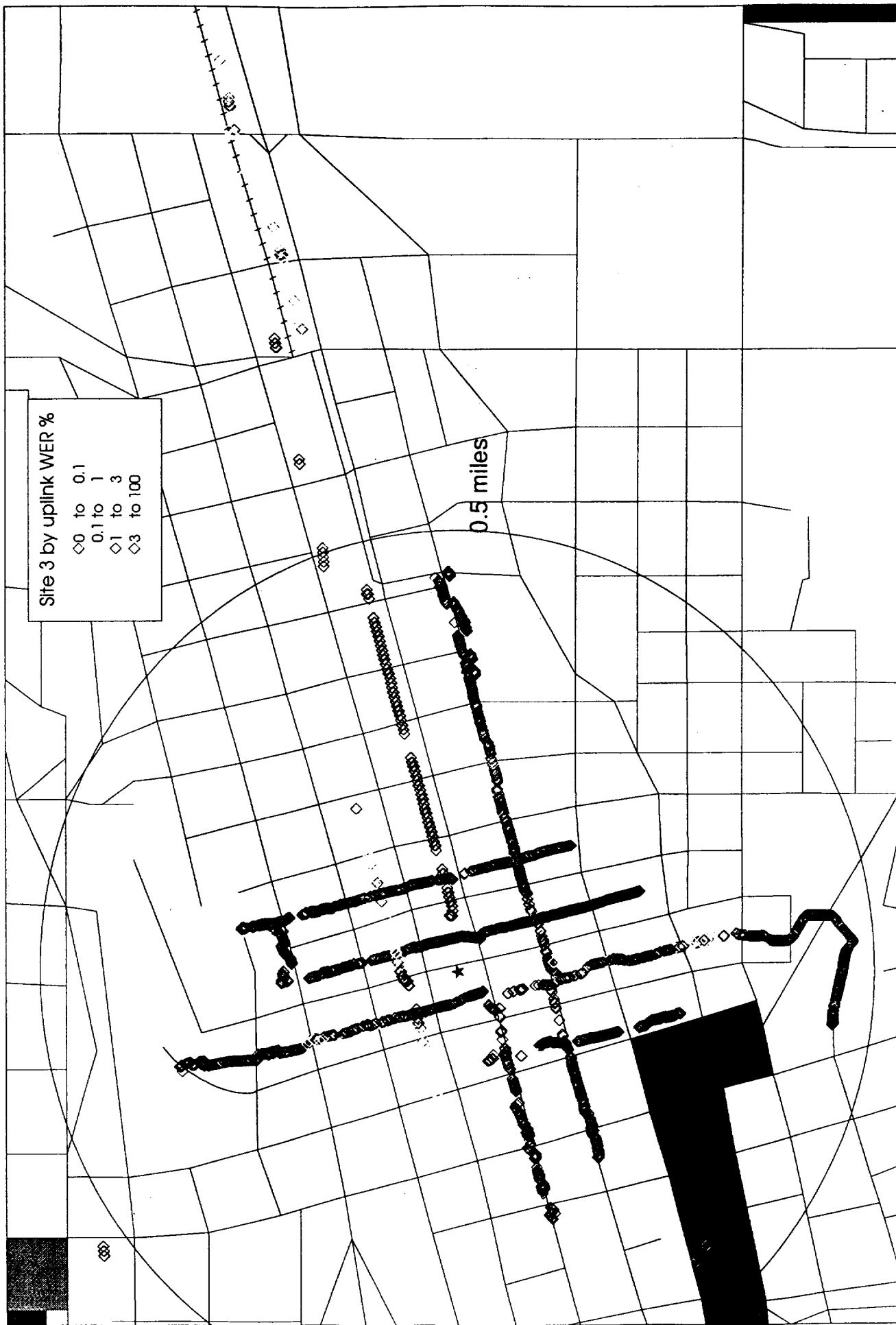
Map 4.2.2.1



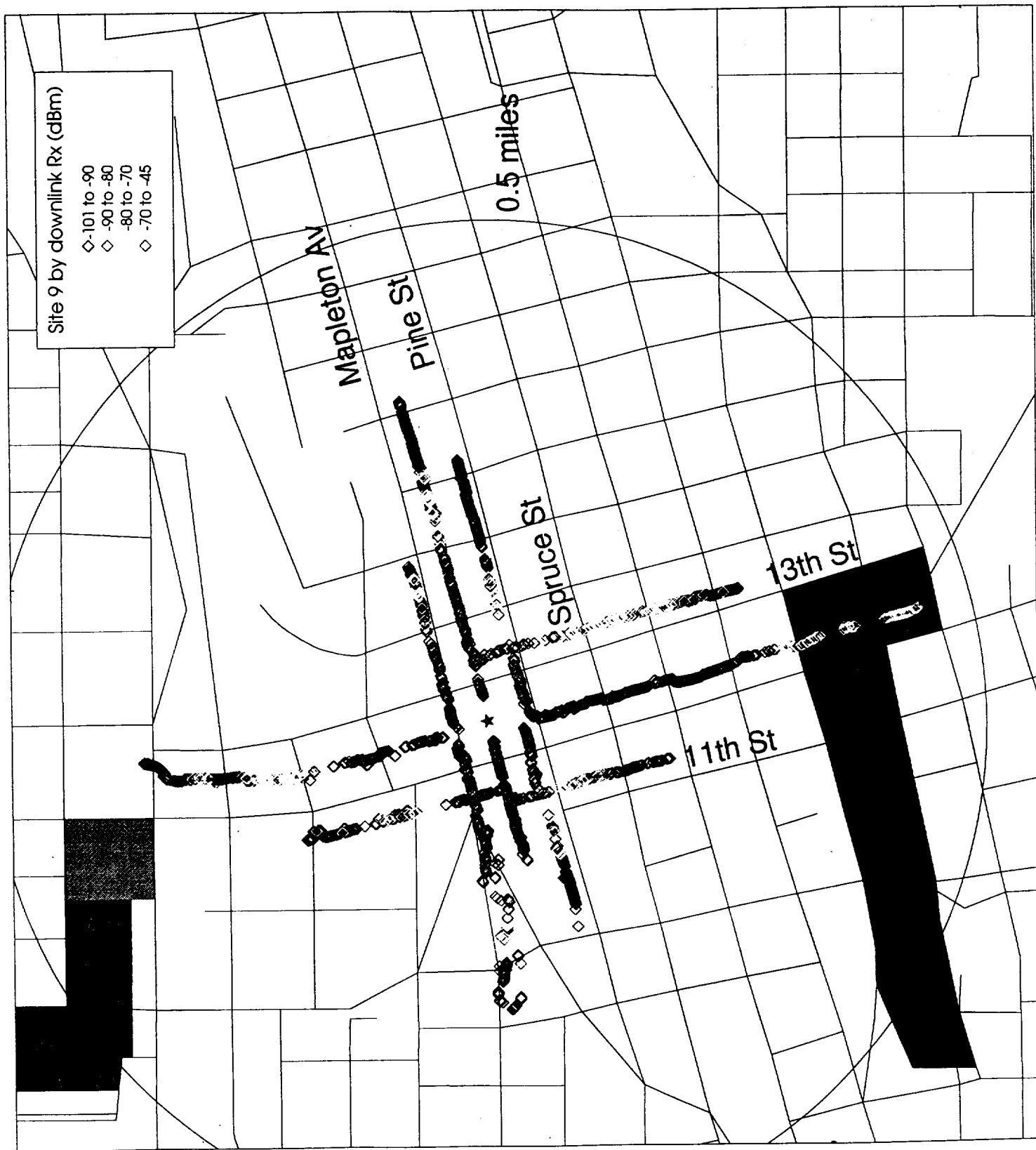
Map 4.2.2.2

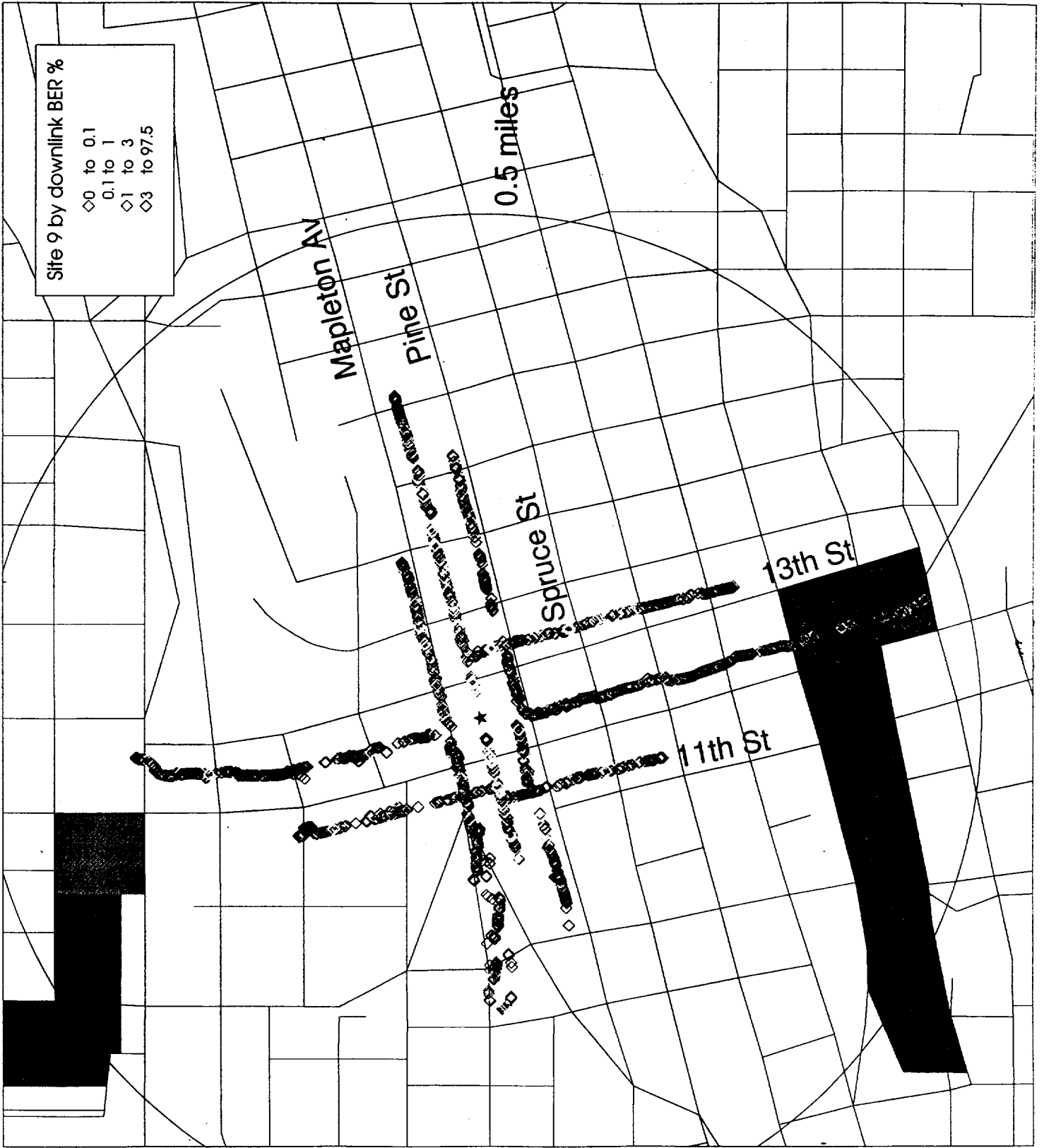


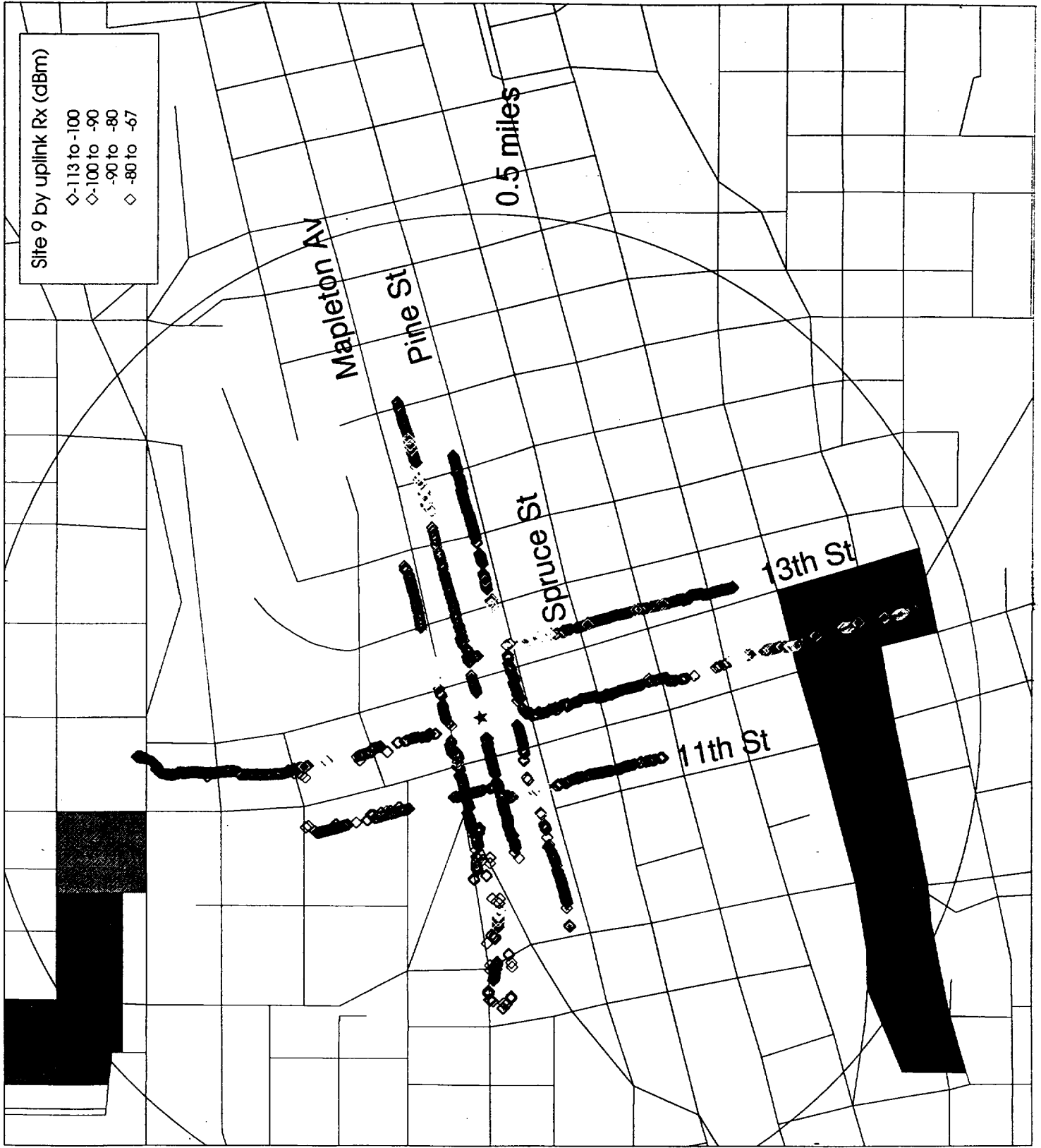
Map 4.2.2.3

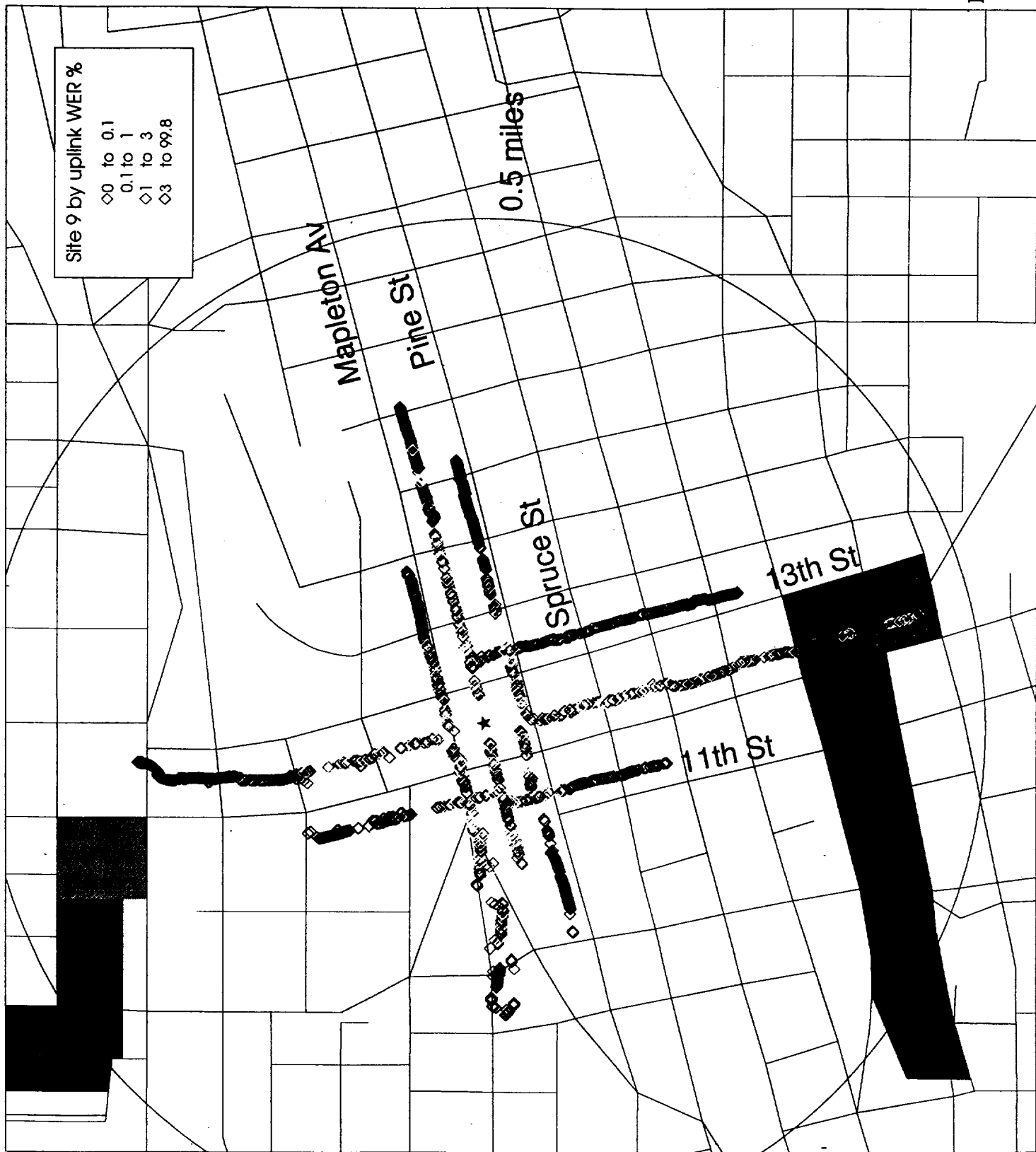


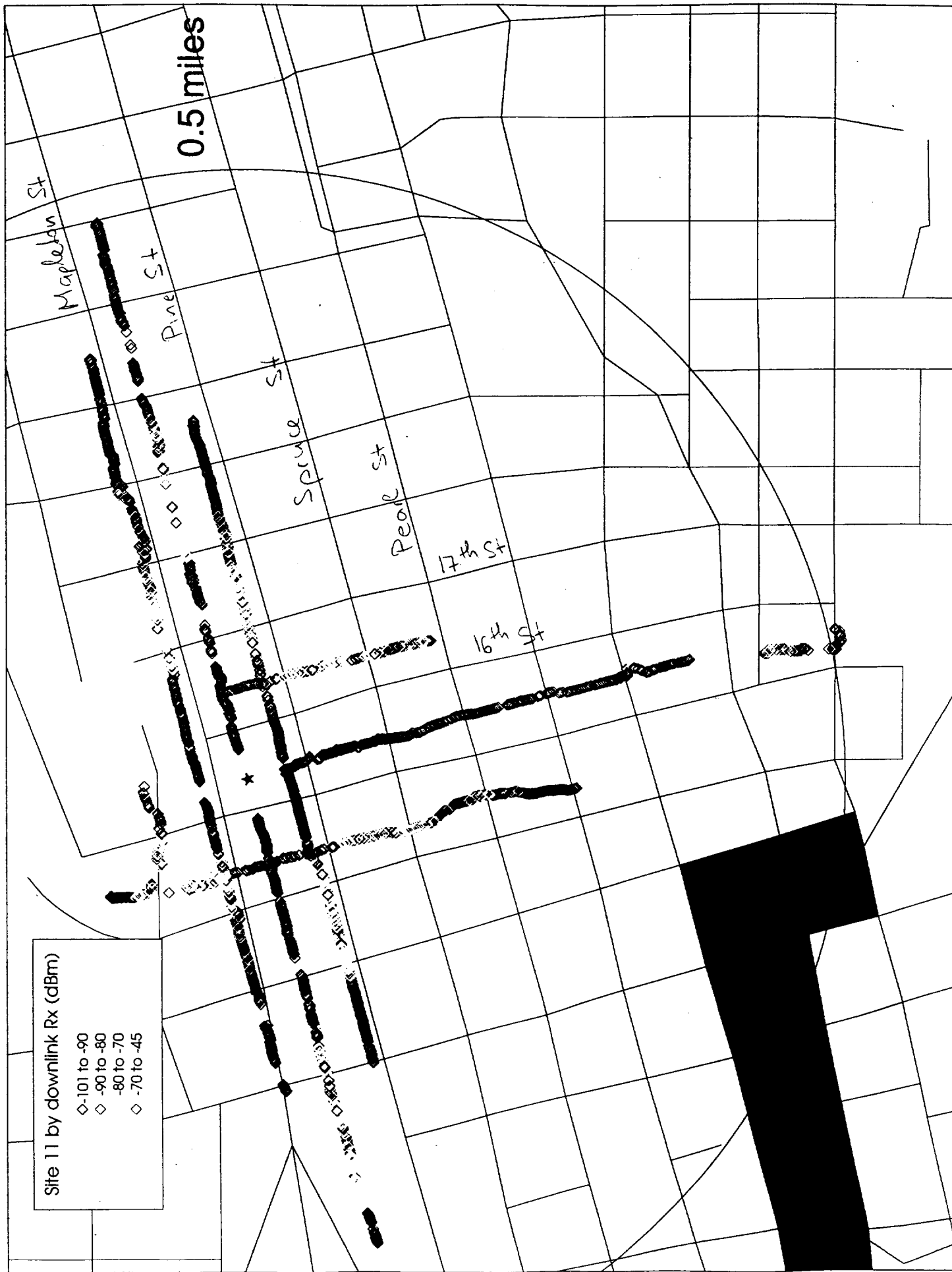
Map 4.2.2.4



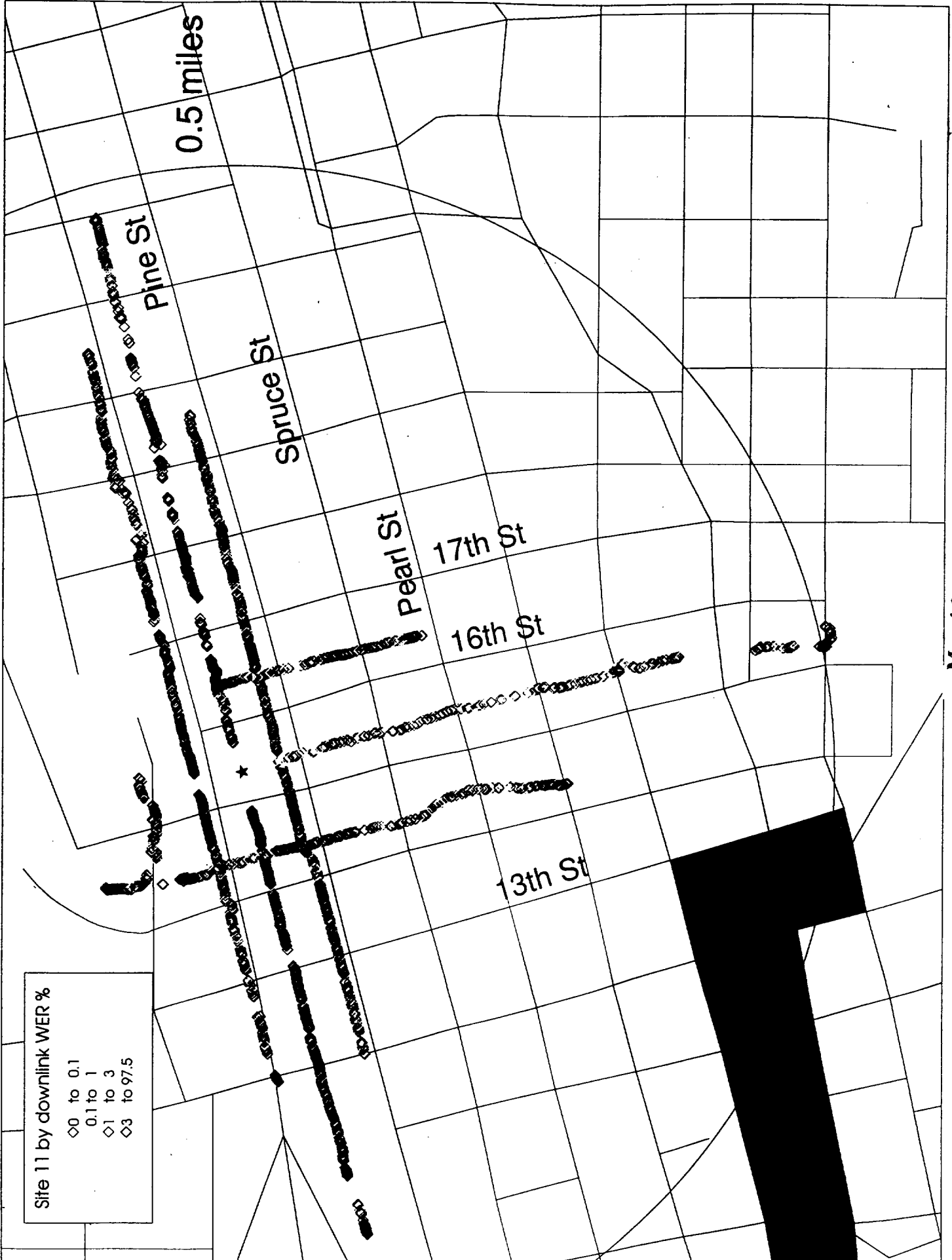




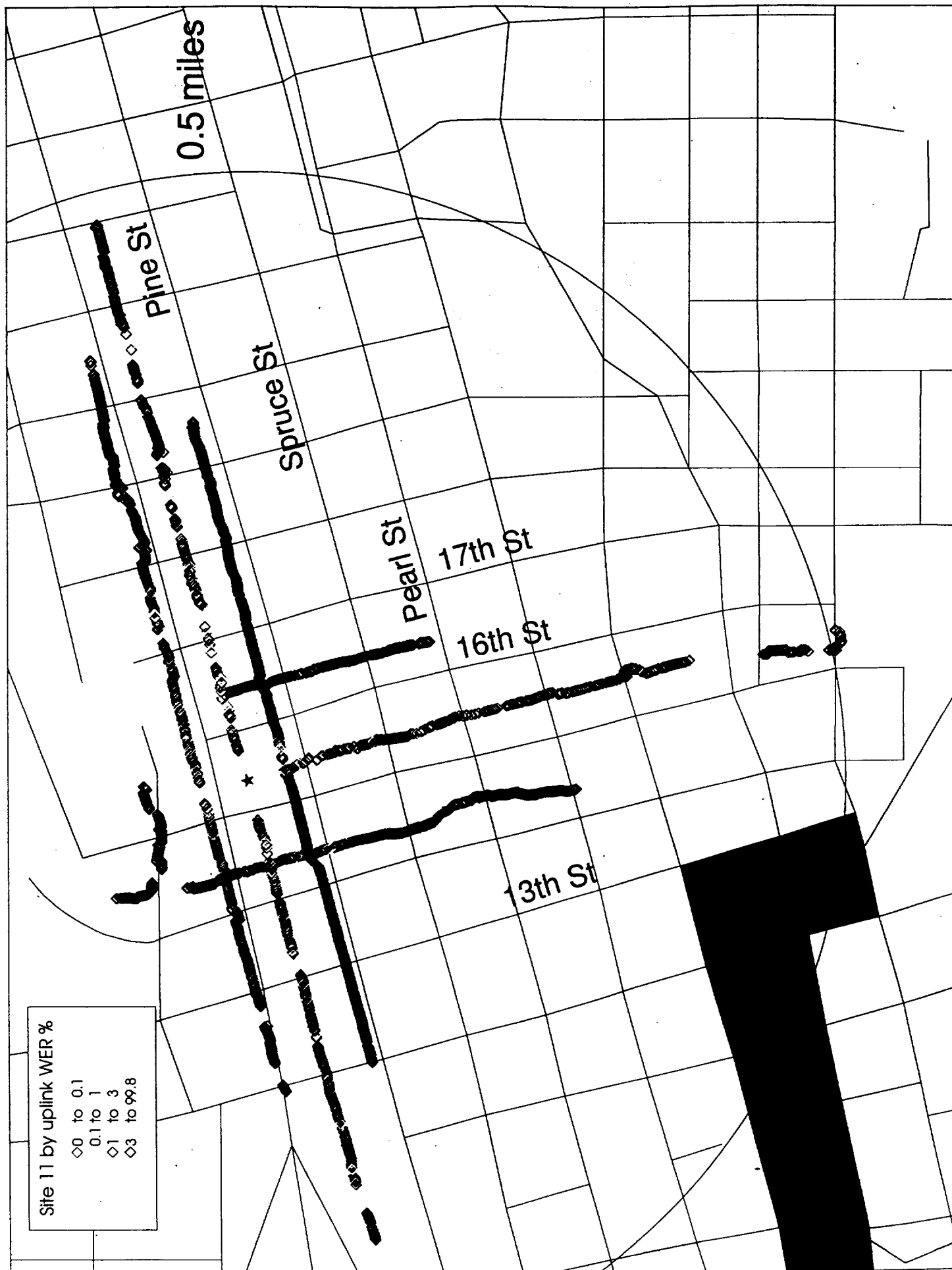


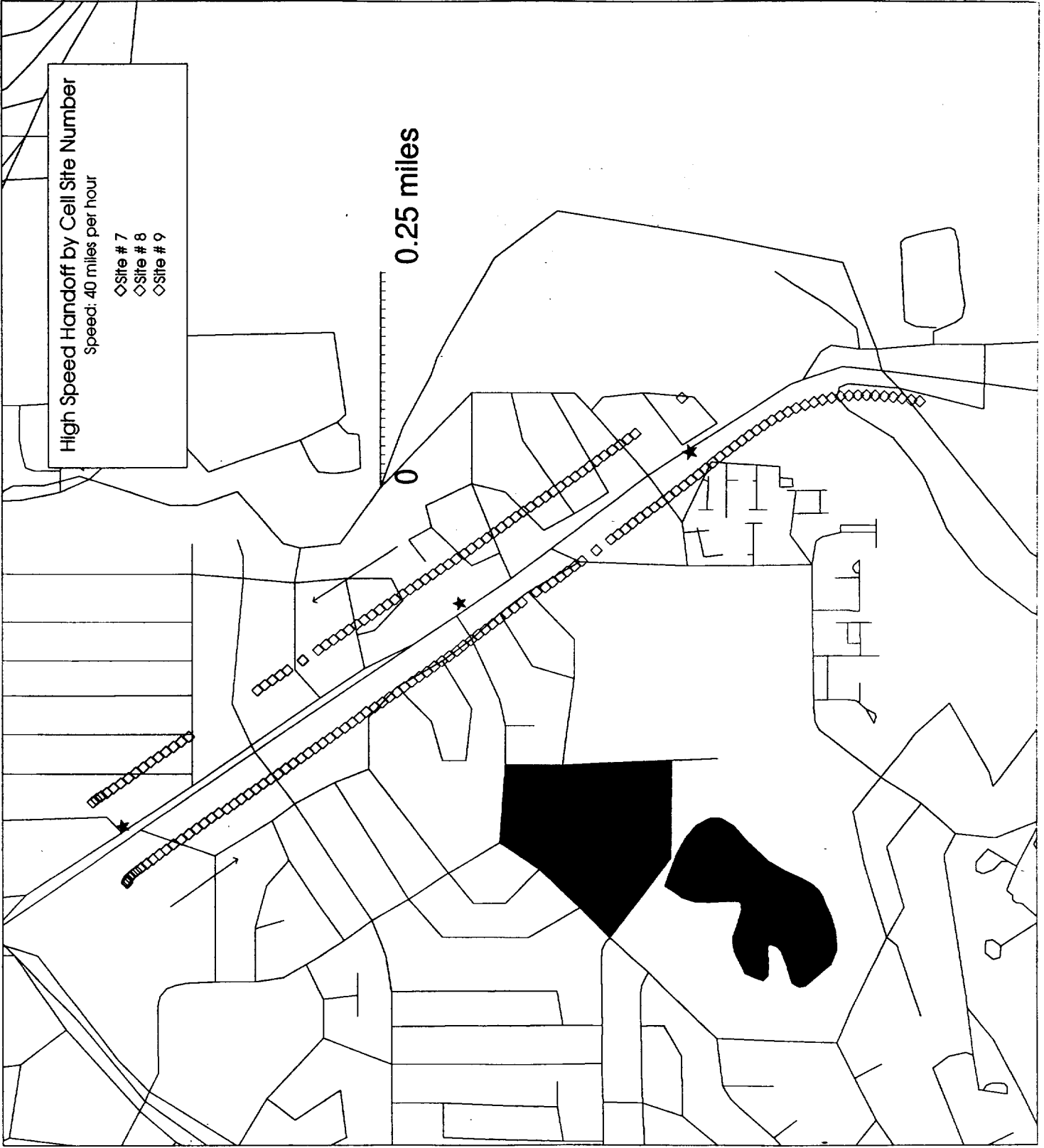


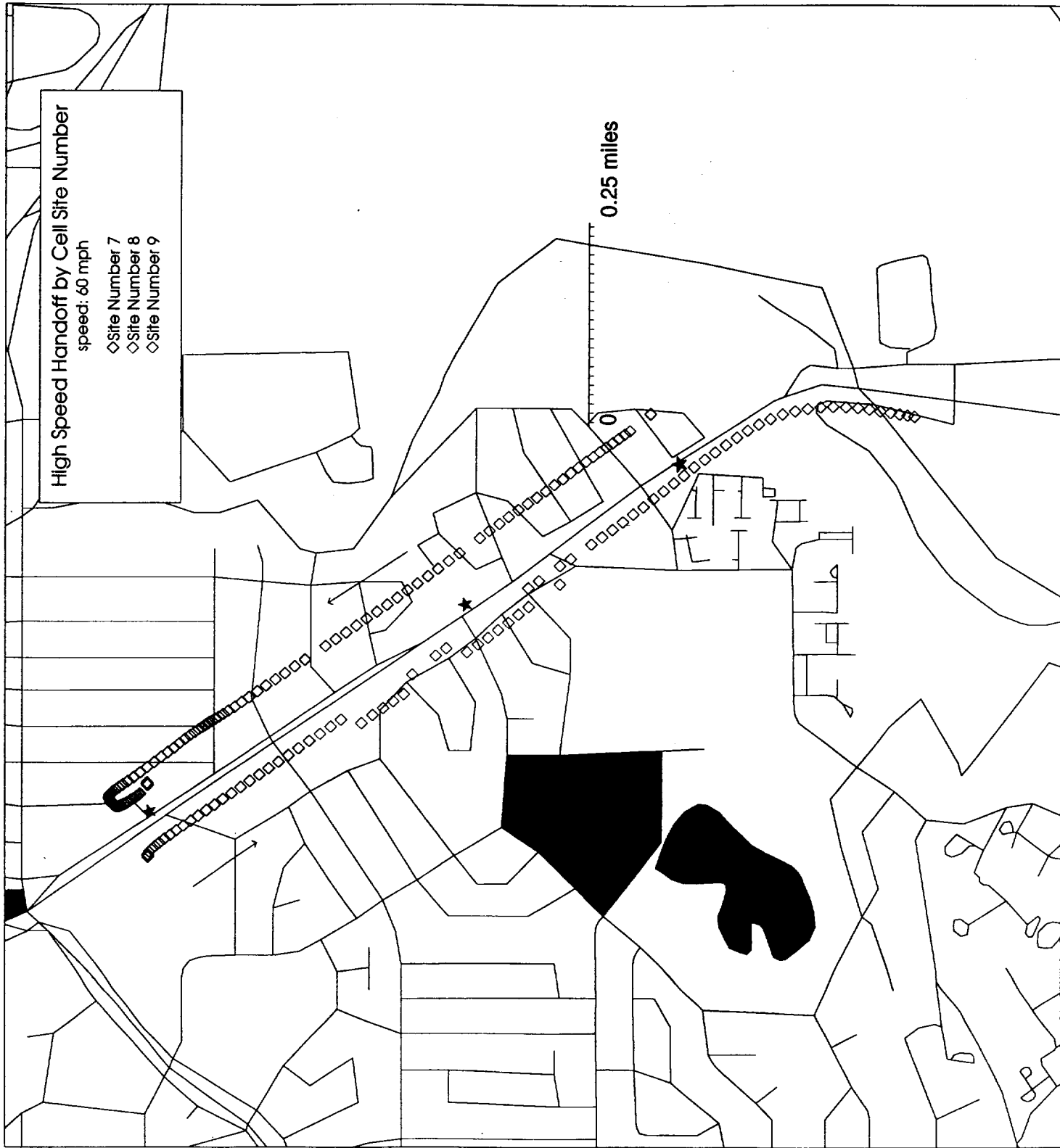
Map 4.2.4.1

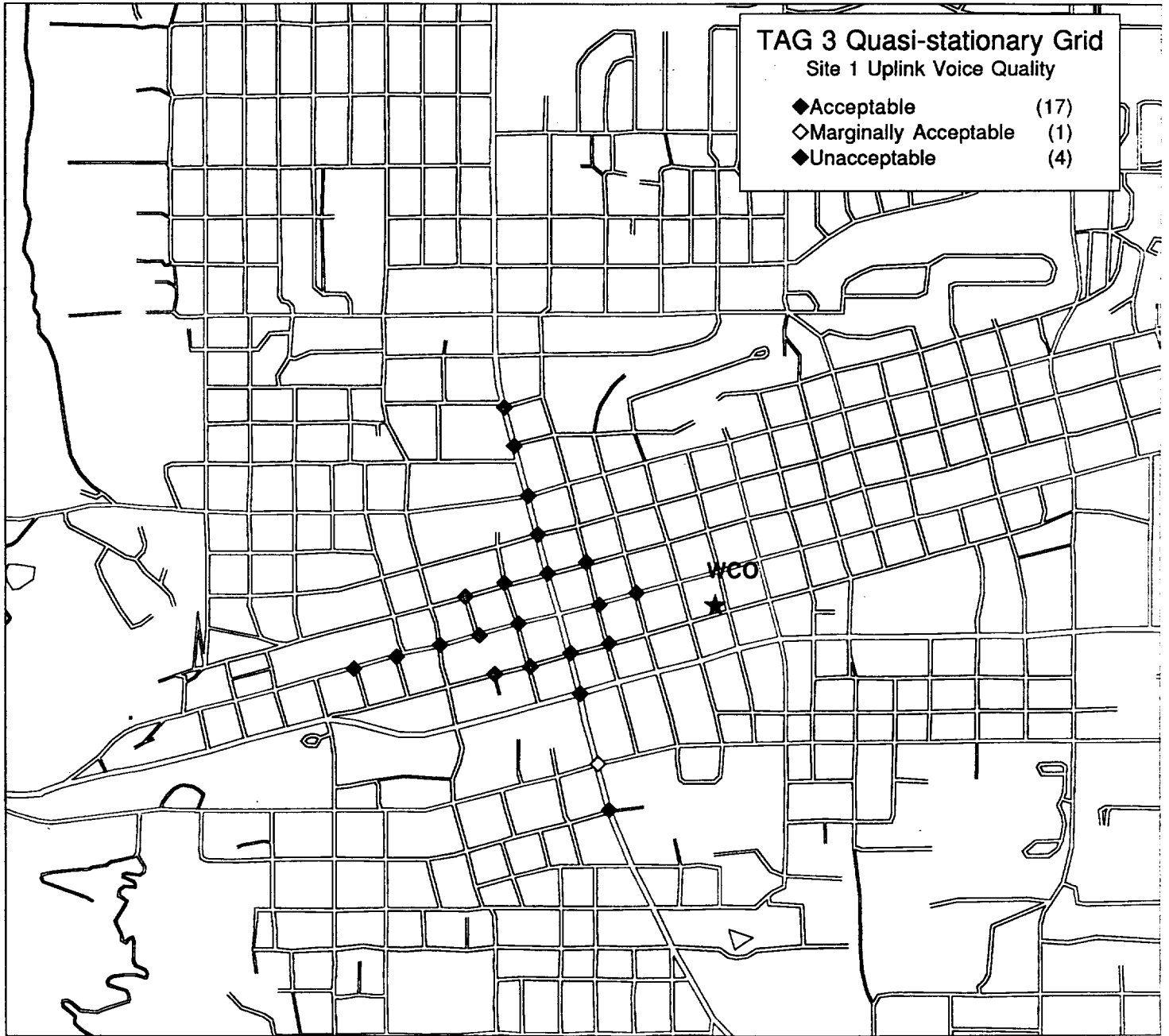


Map 4.2.4.2

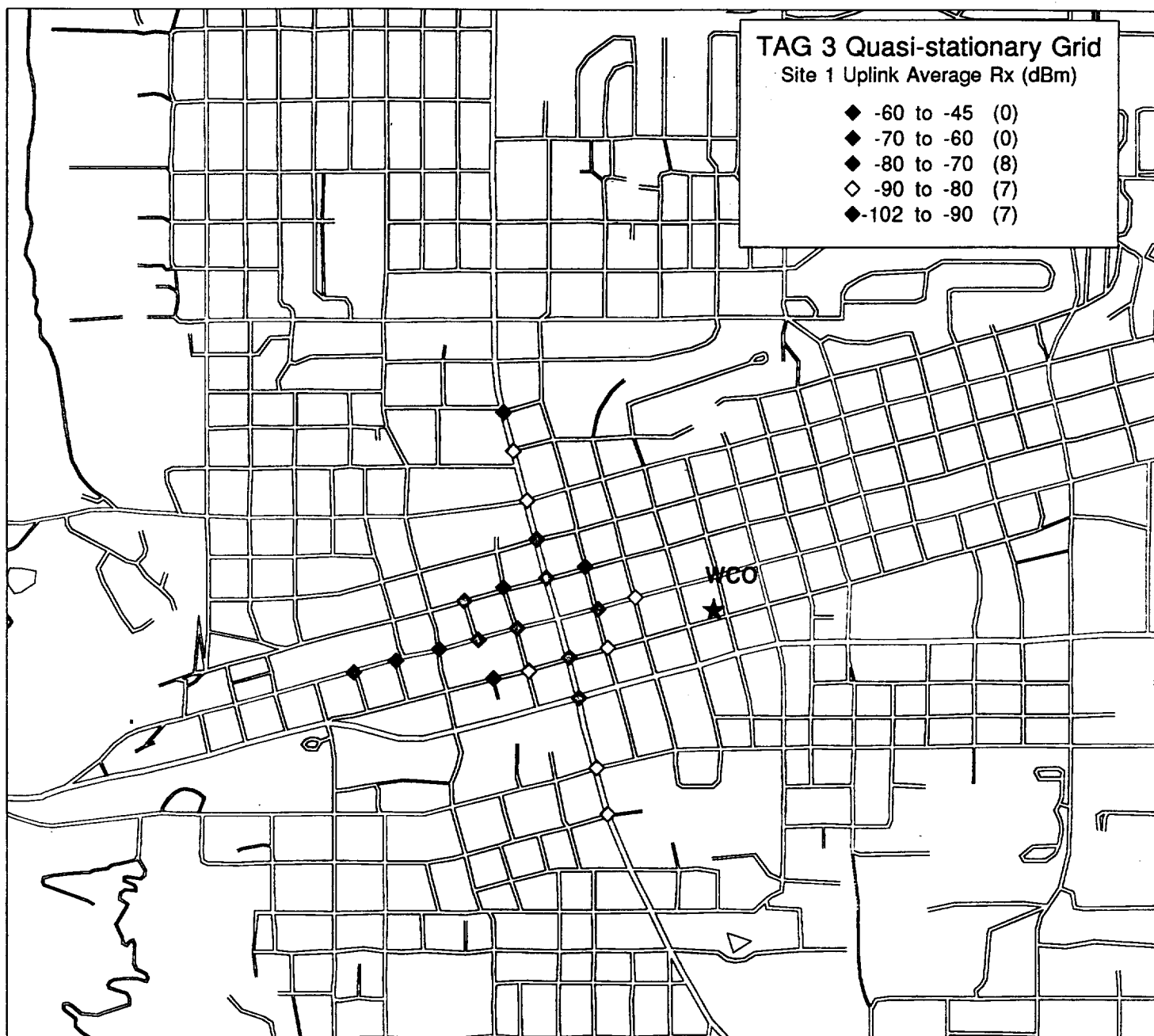




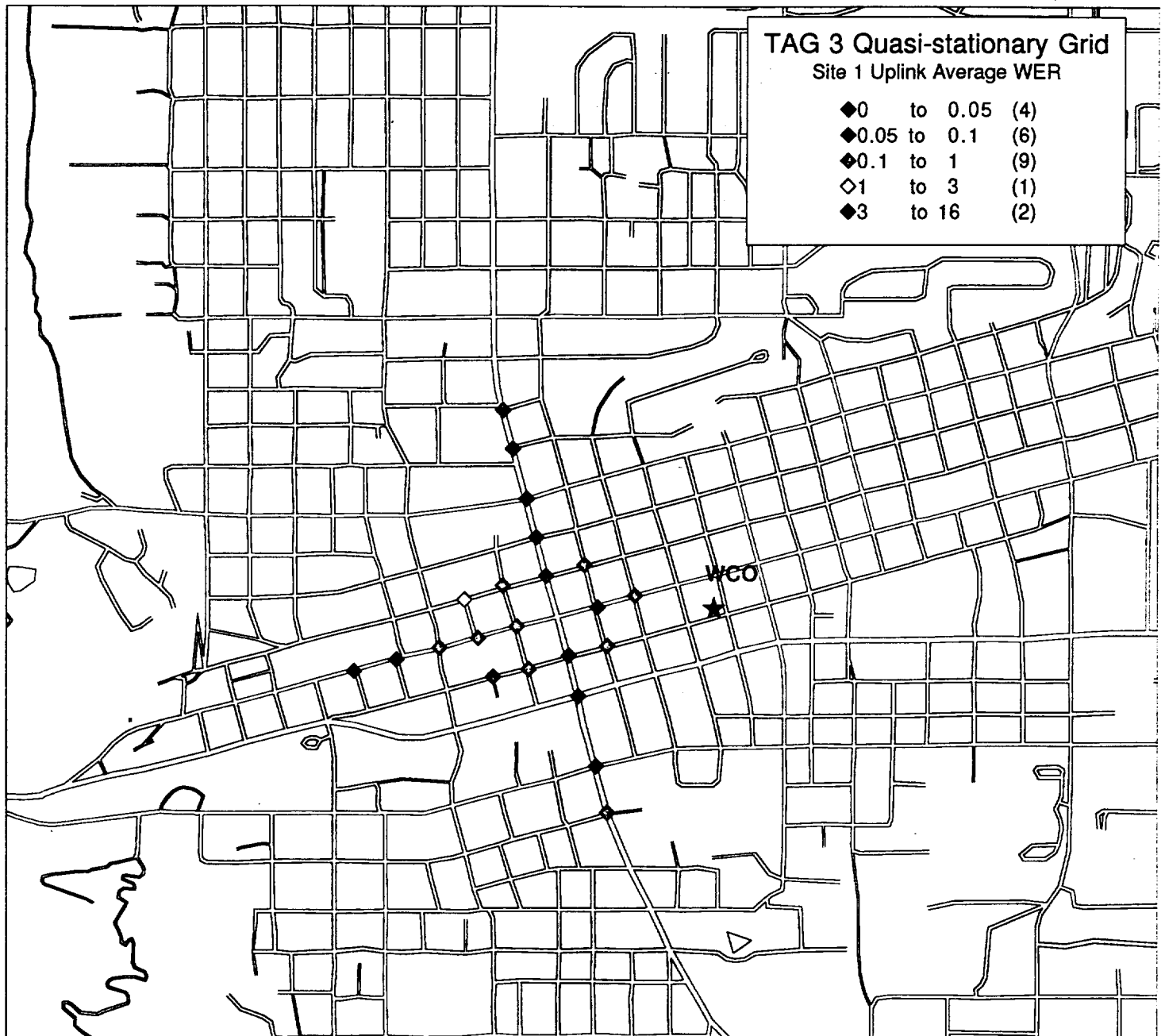




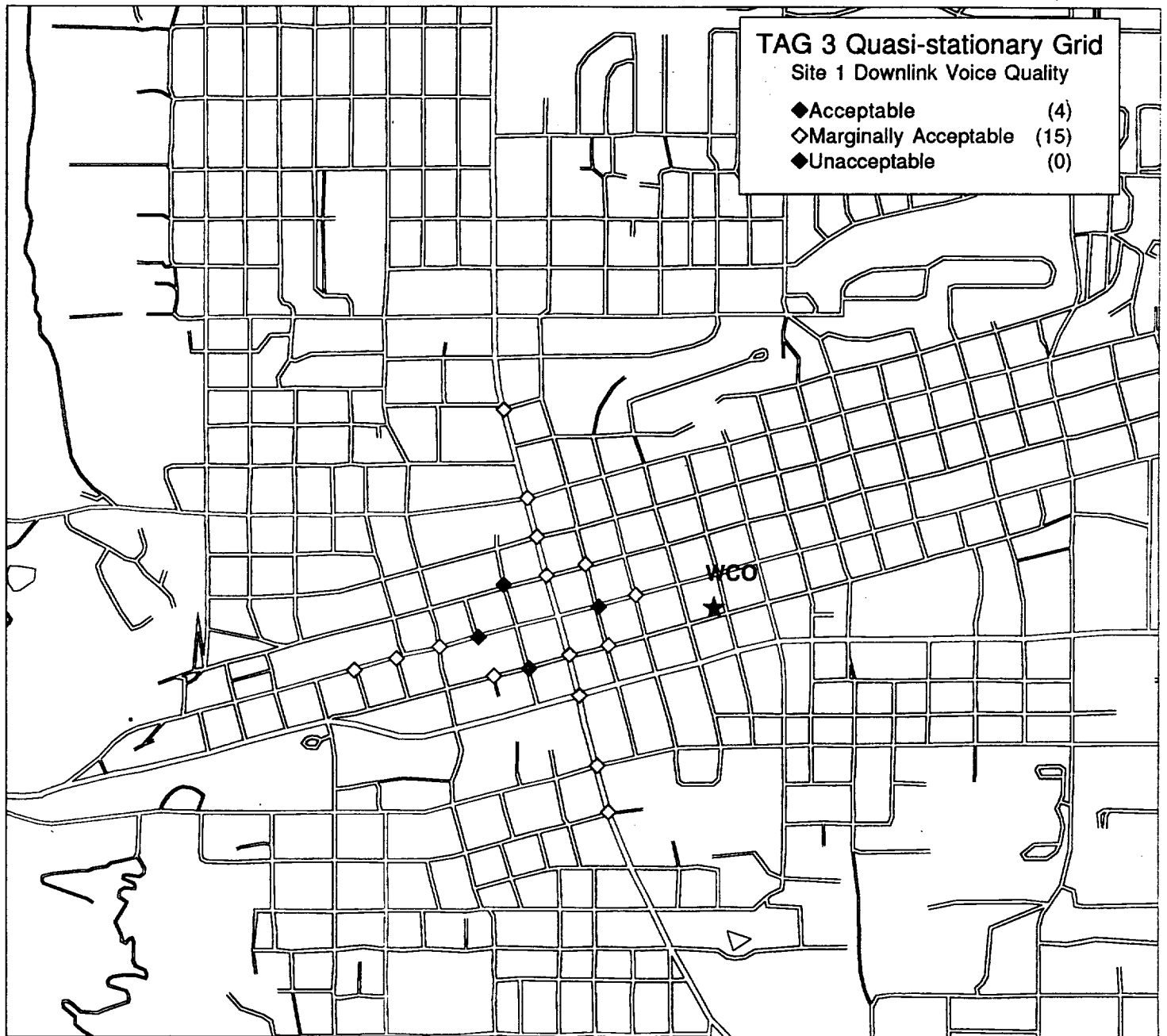
Map 7.3.5.1



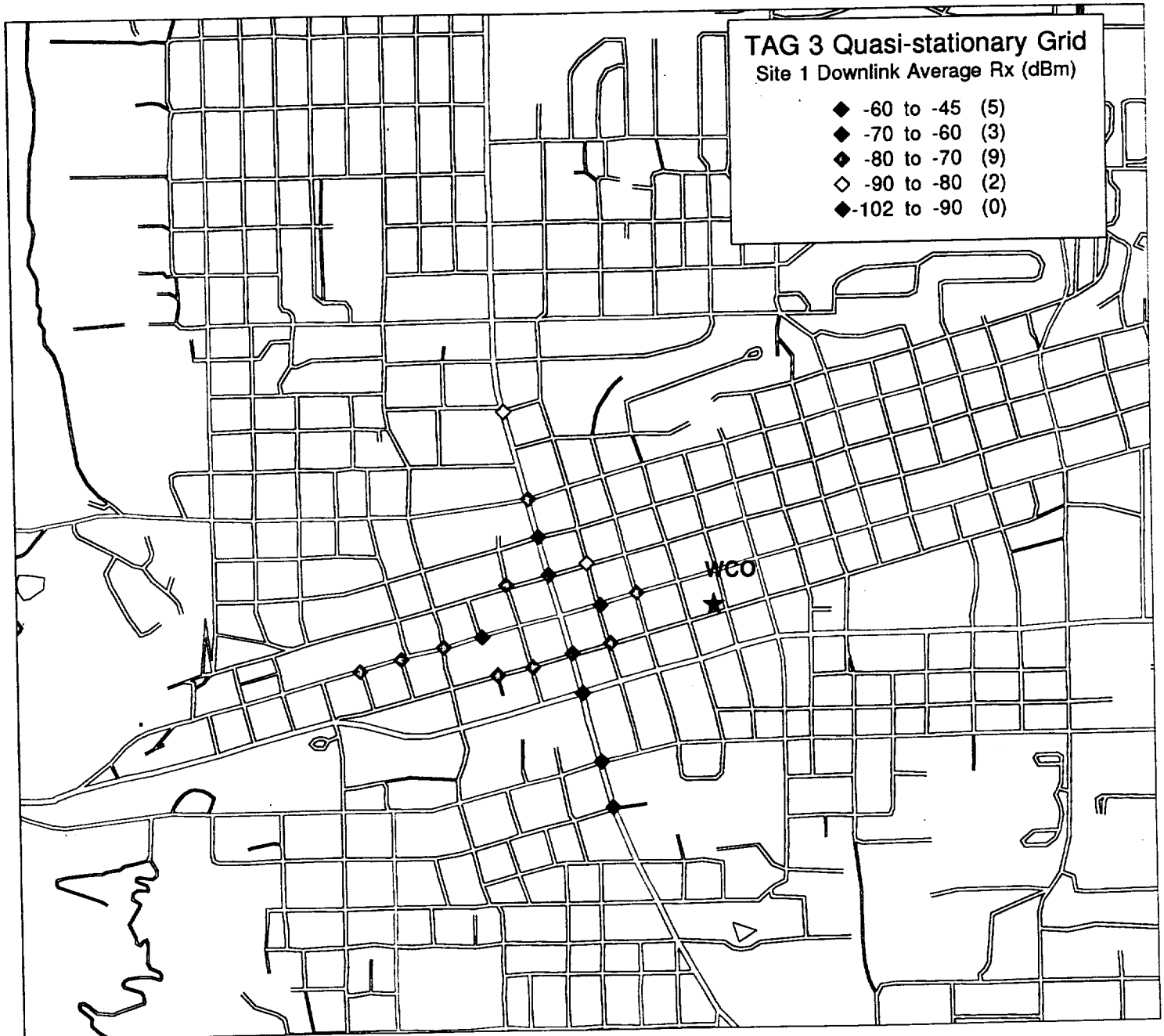
Map 7.3.5.2



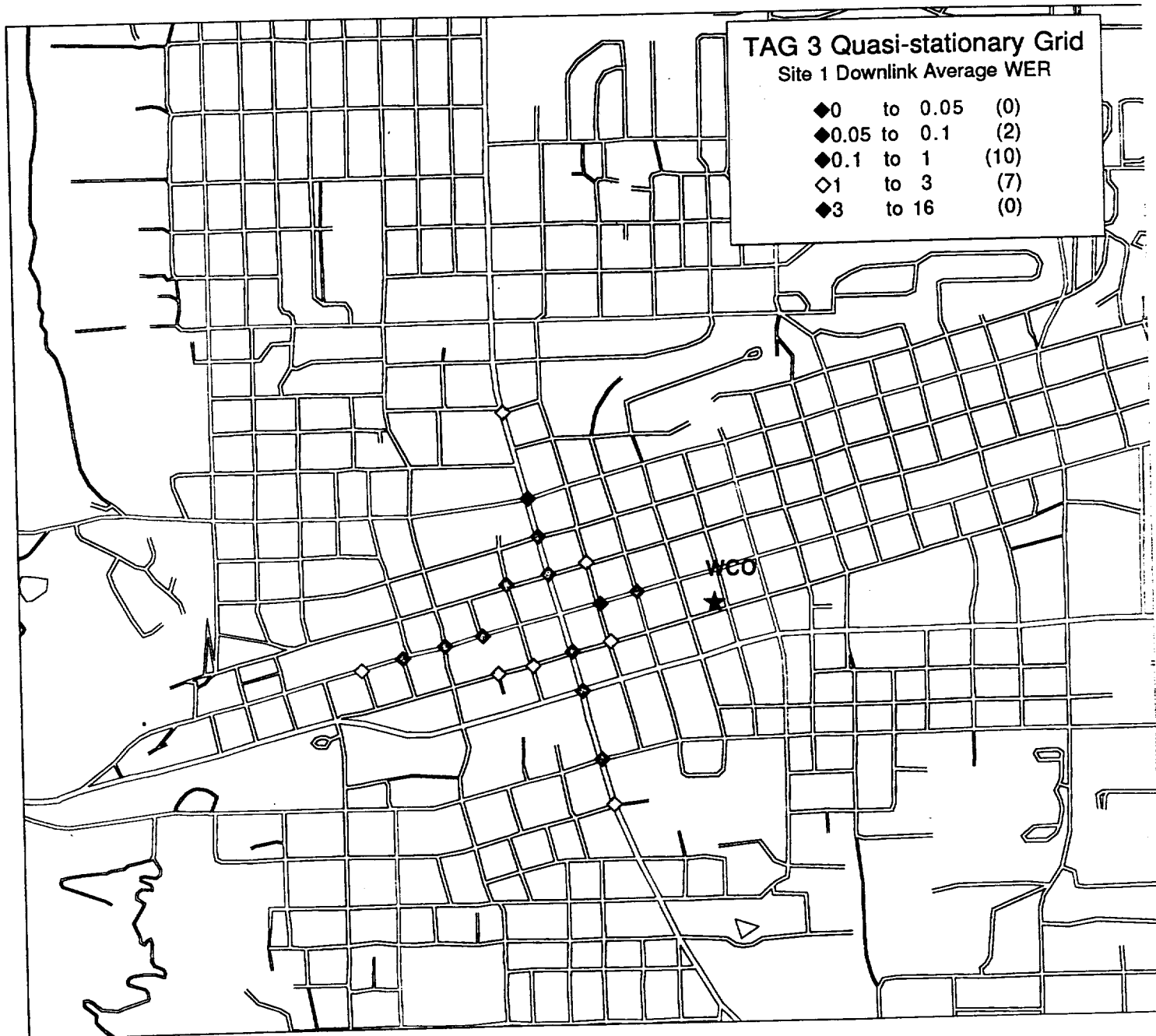
Map 7.3.5.3



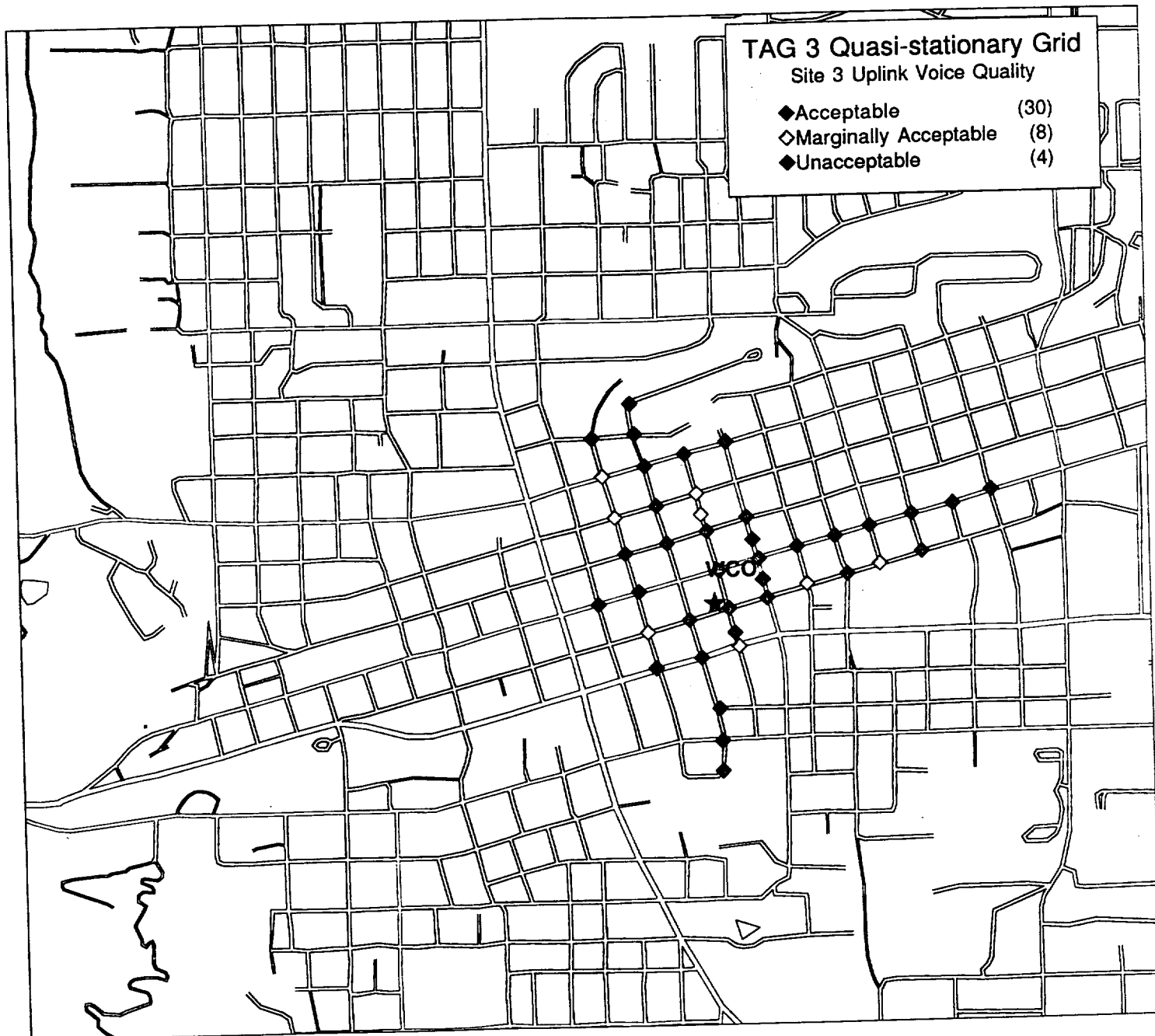
Map 7.3.5.4



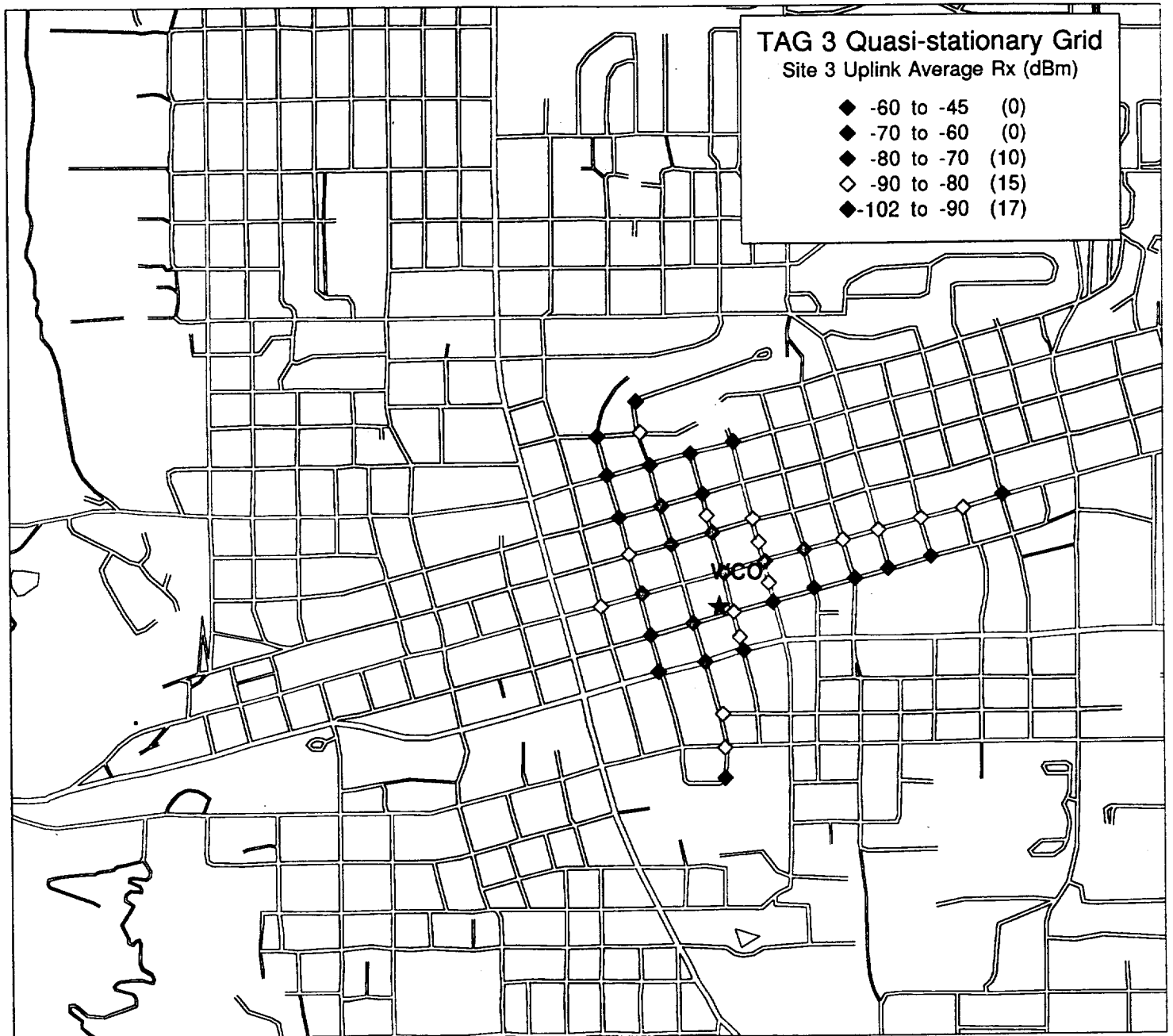
Map 7.3.5.5



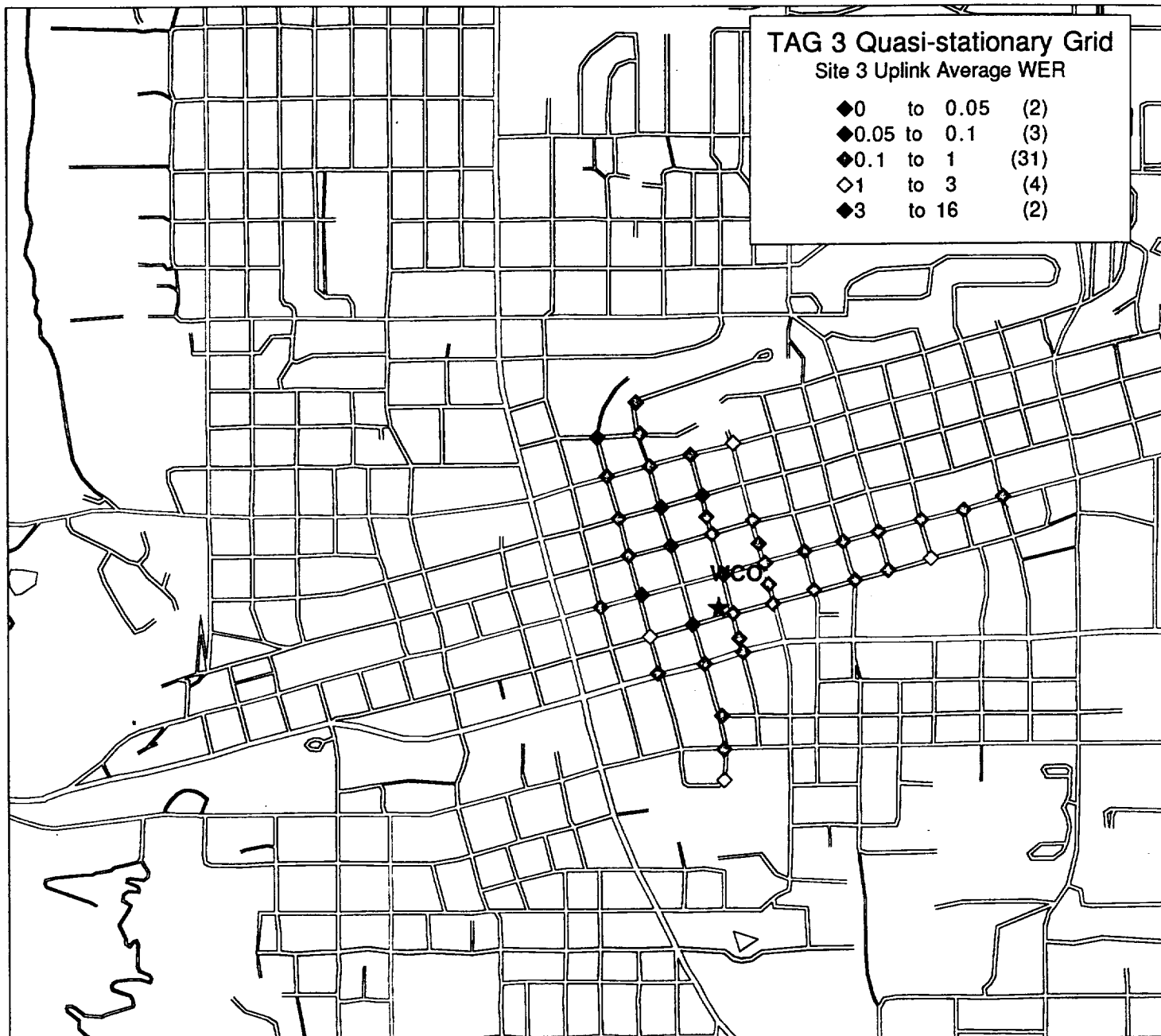
Map 7.3.5.6



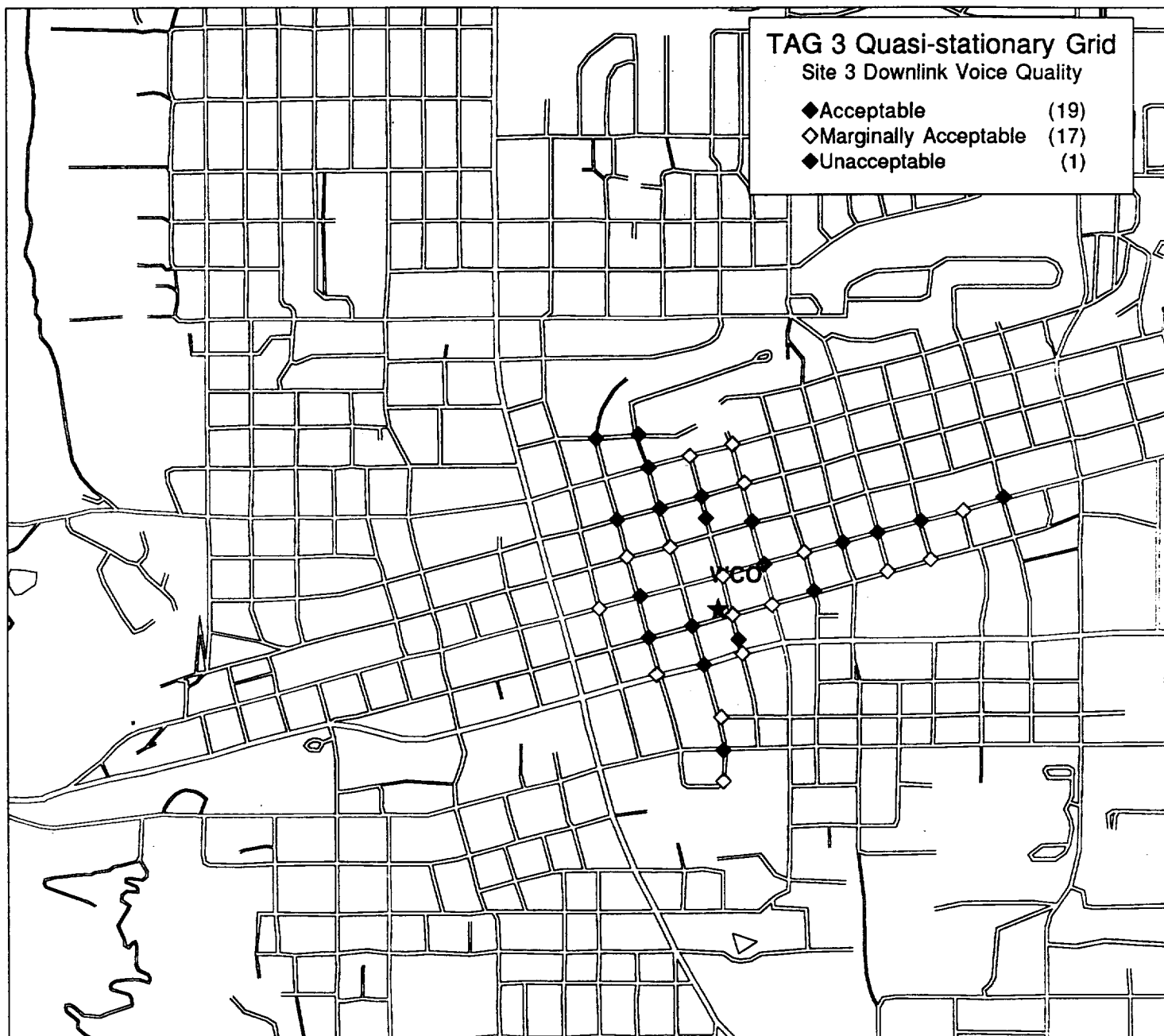
Map 7.3.5.7



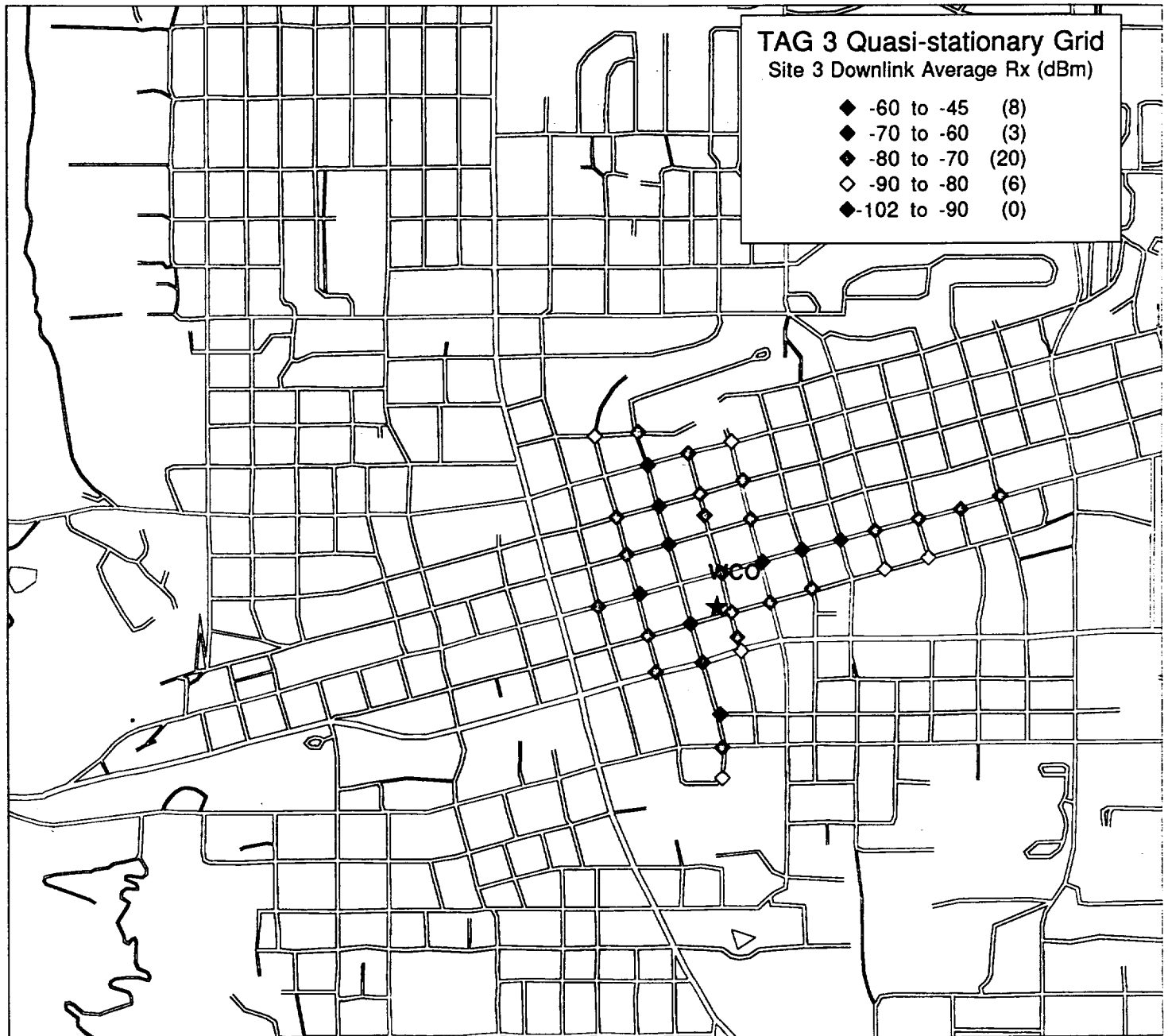
Map 7.3.5.8



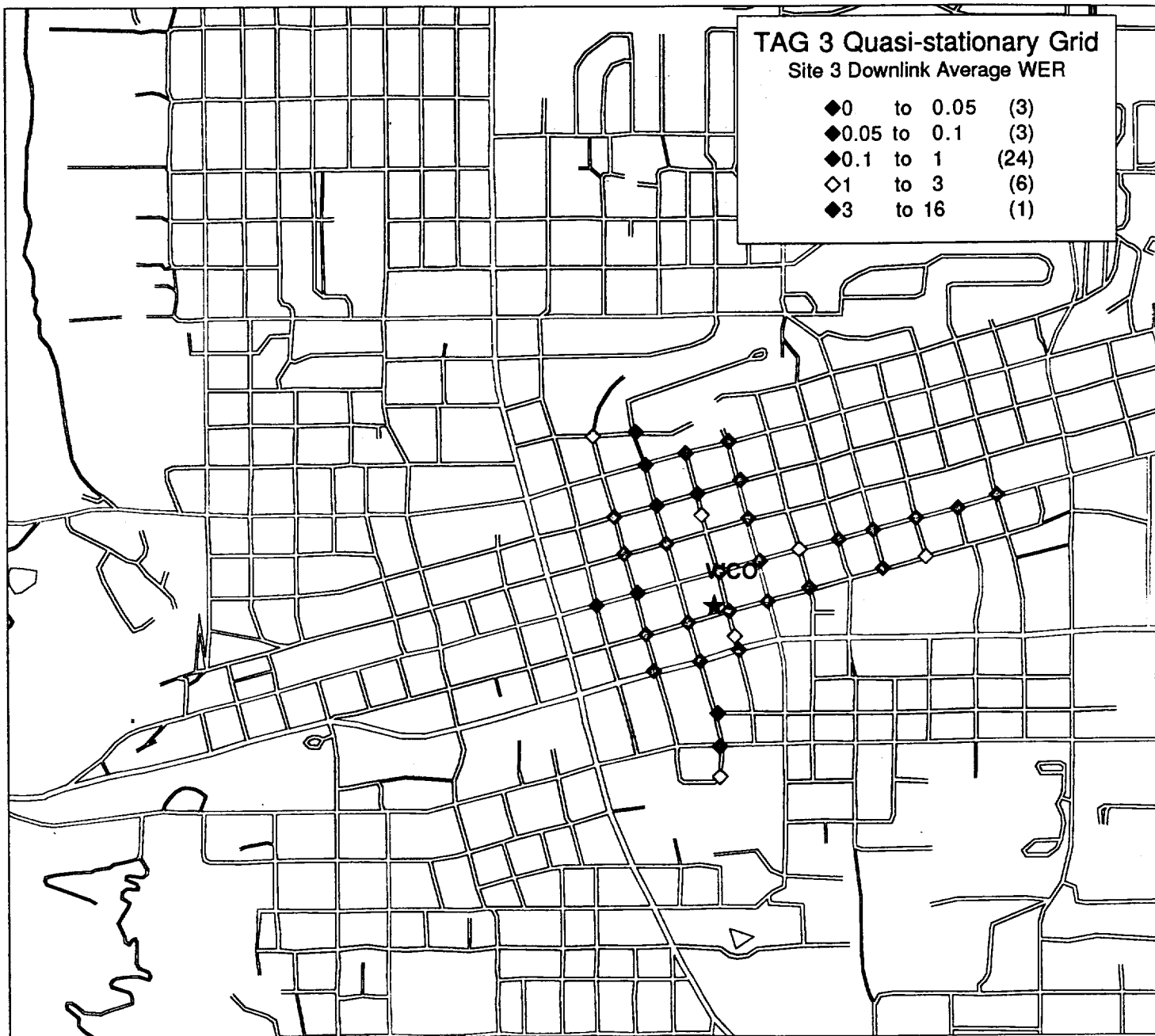
Map 7.3.5.9



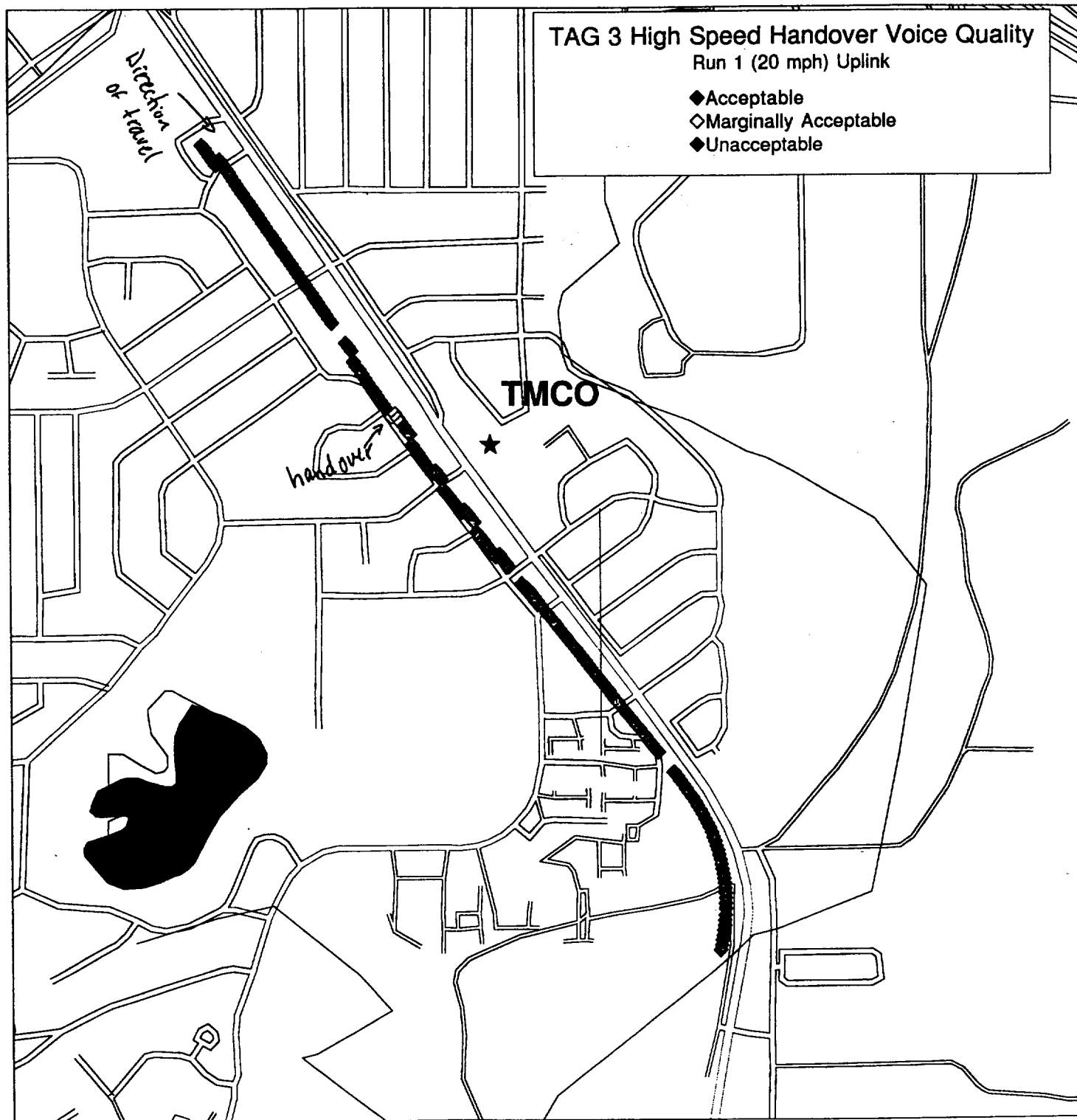
Map 7.3.5.10



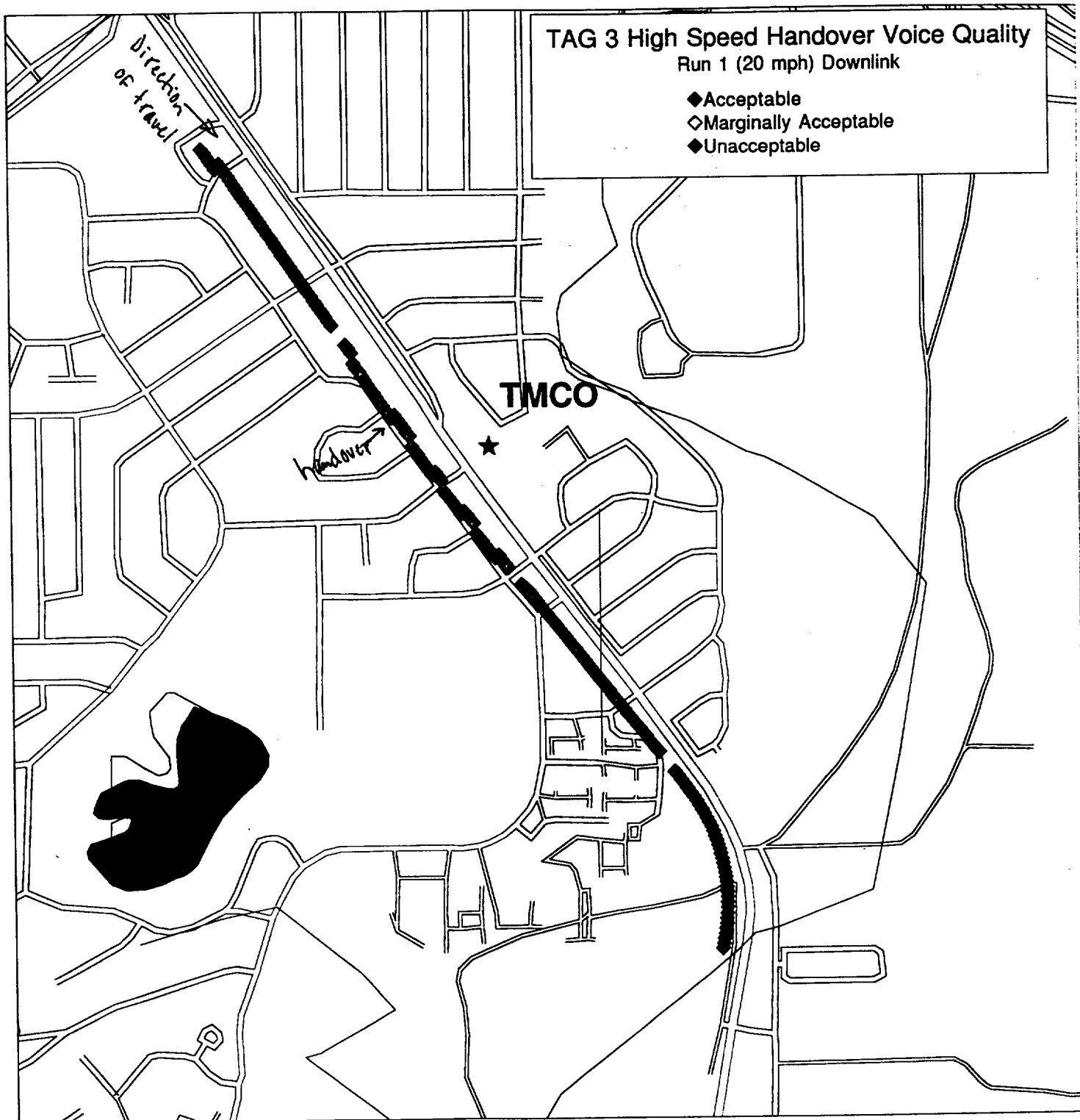
Map 7.3.5.11



Map 7.3.5.12



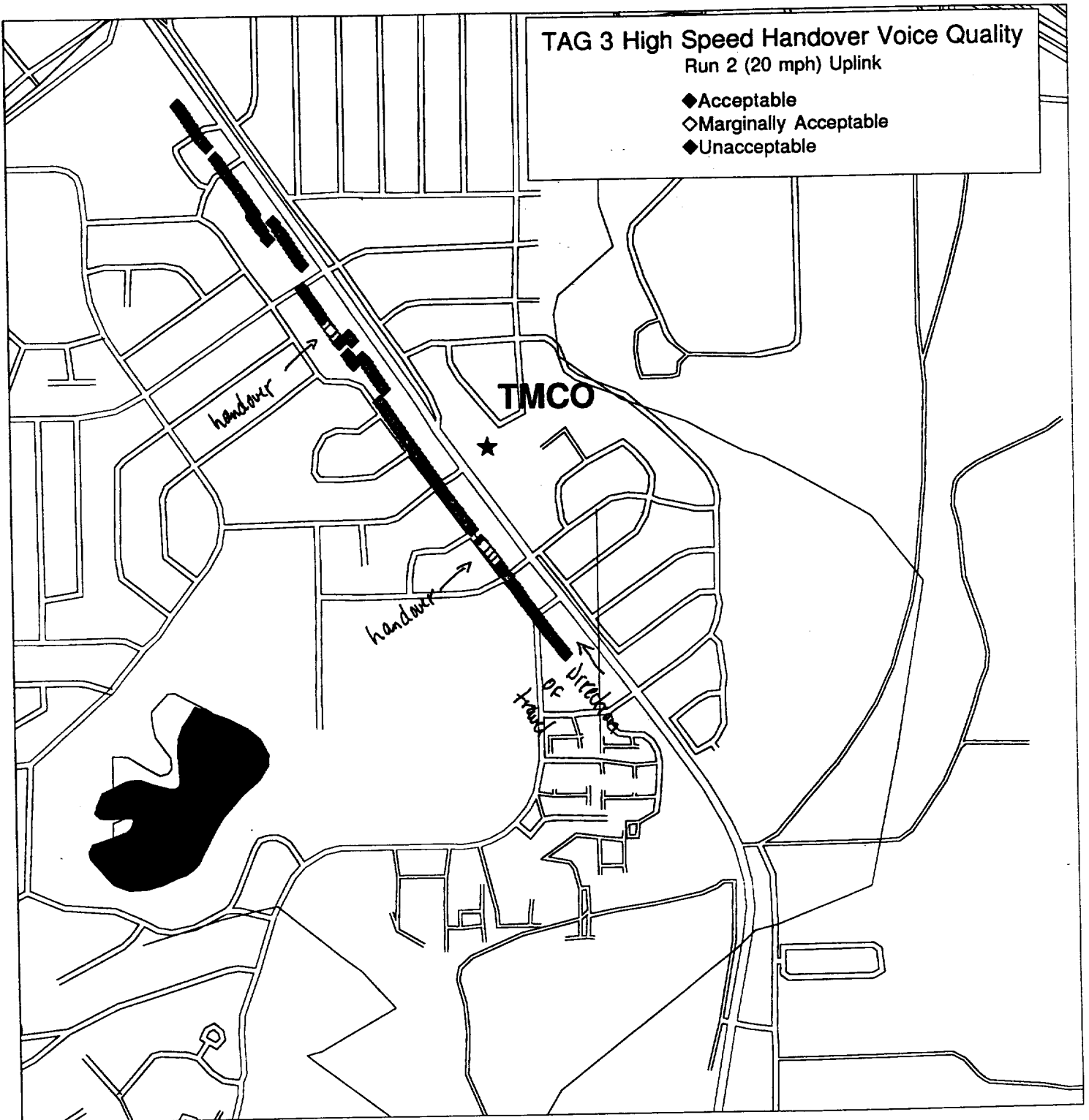
Map 7.3.6.1



Map 7.3.6.2

TAG 3 High Speed Handover Voice Quality
Run 2 (20 mph) Uplink

- ◆ Acceptable
- ◇ Marginally Acceptable
- ◆ Unacceptable

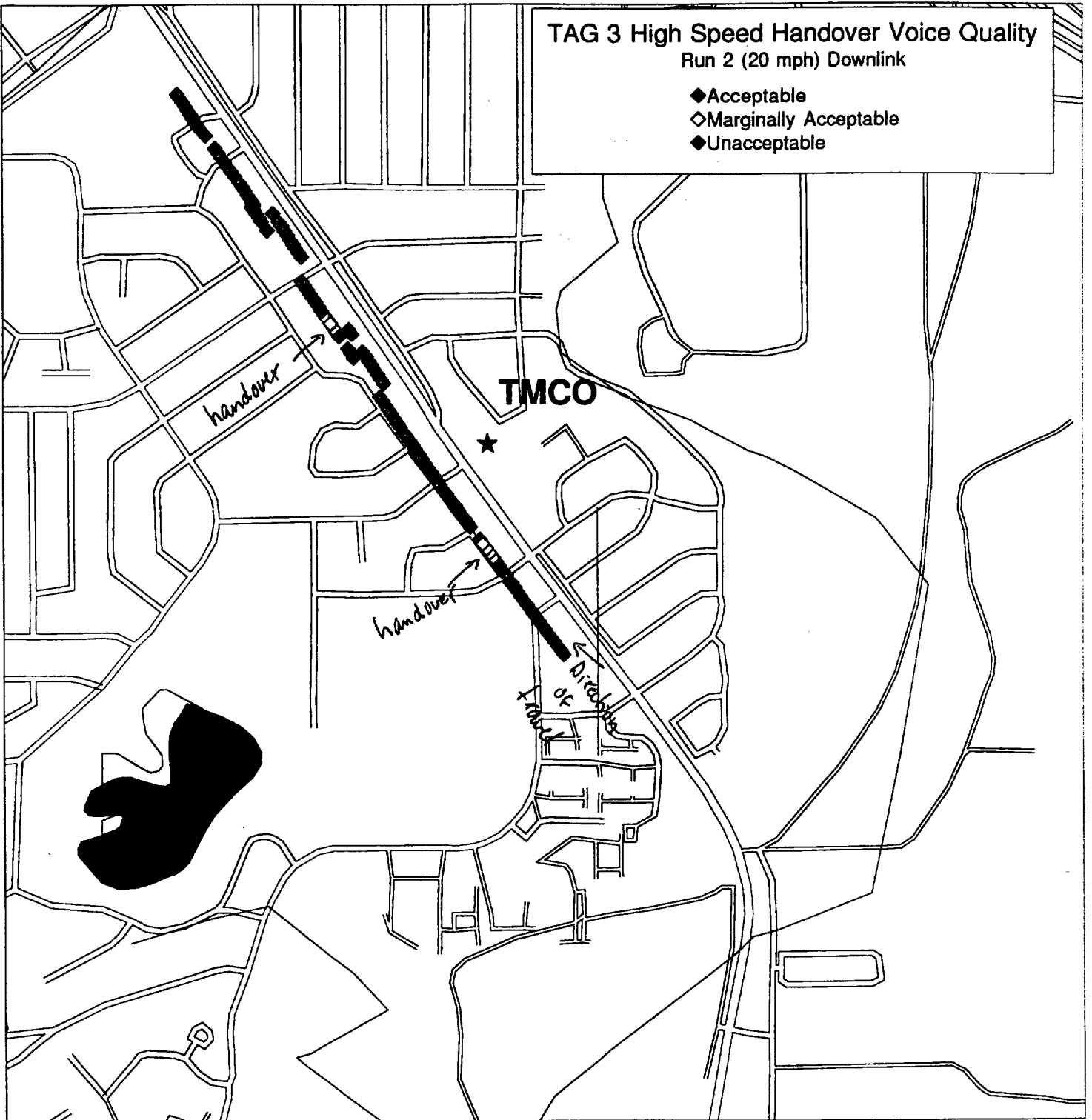


Map 7.3.6.3

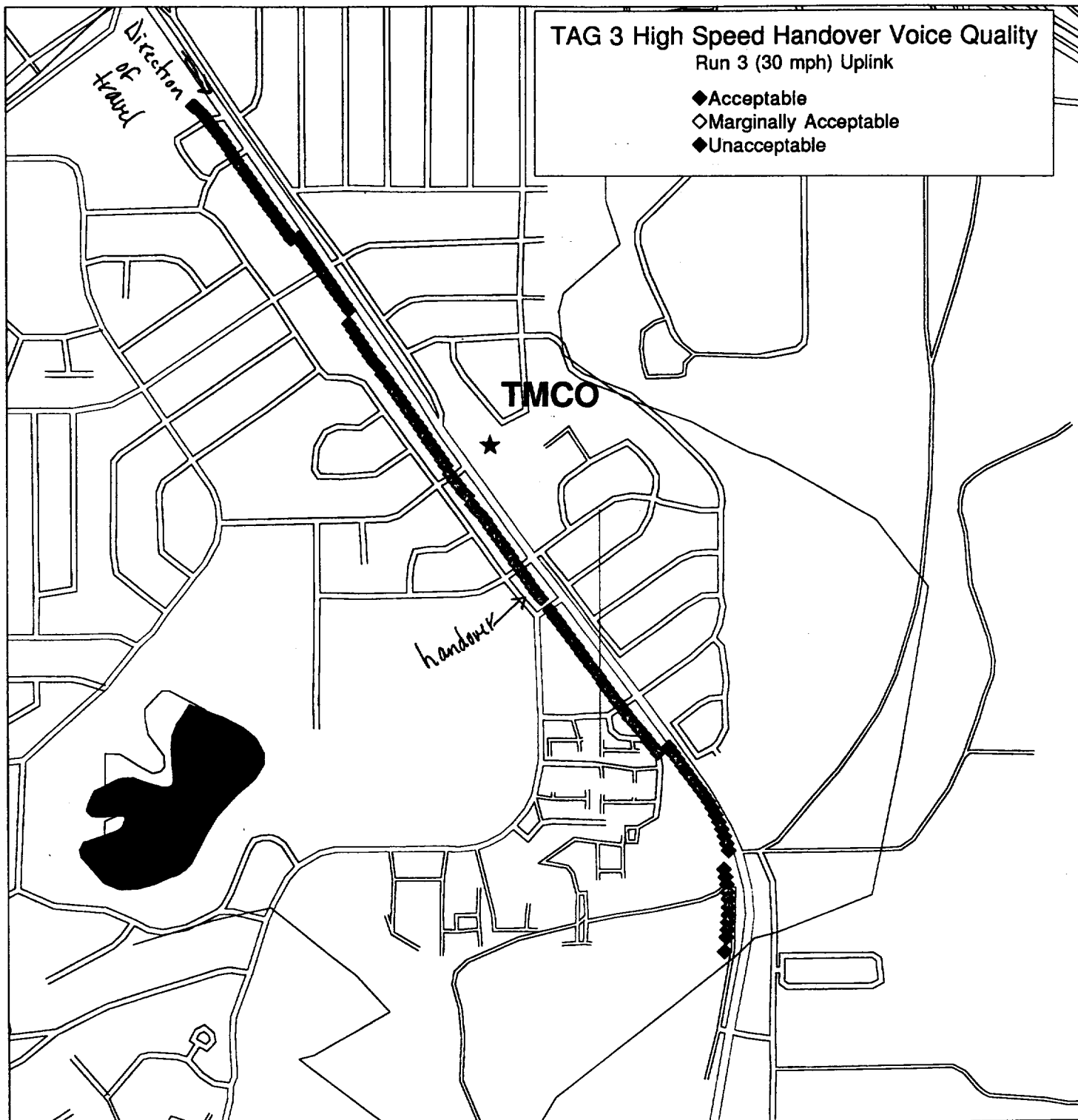
TAG 3 High Speed Handover Voice Quality

Run 2 (20 mph) Downlink

- ◆ Acceptable
- ◇ Marginally Acceptable
- ◆ Unacceptable



Map 7.3.6.4



Map 7.3.6.5