Recognition of the heterogeneity of patients with aphasia, apraxia of speech and other acquired adult speech and language disorders has given impetus to single subject treatment efficacy research that utilizes time-series analysis or a multiple-baseline design. Statistical methods used in this line of inquiry differ from those employed in group based studies and have evolved over the last two decades. In the early 80-ties researchers (Tryon, 1982; Salzberg, 1987) proposed the *C*-statistic as a more appropriate alternative to Cohen's *d* statistic (see Figure 1.) for time-series data. The *d*-statistic is traditionally used in group based research to analyze pre- versus post-treatment data.

Unlike Cohen's *d*, the *C*- statistic is derived from single-subject data that can be limited to as few as eight observations. Considering that single-subject studies often contain a limited number of repeat observations, this characteristic was desirable rendering it the technique of choice in a number of single subject designs in acquired communication disorders (Pring, 1986; Ballard, Barlow, & Robin, 2001). Concerns raised by Crosbie (1989; 1993)that the *C*-statistic lacked sensitivity in measuring treatment slope led Robey (1998a) to recommended the use of interrupted time-series analysis correlations or ITSACORR (Salzberg, 1987) for single-subject analysis in the field of speech-language pathology. However, even as several investigators in speech pathology adopted the technique in their studies (Kendall, Rodriguez, Rosenbek, Conway & Gonzales-Rothi, 2006; Kearns, 2007), Huitema & McKean (2000a; 2000b) and Huitema (2004) found that the ITSACORR method was still problematic where type 1 error rates were concerned. In view of that finding Huitema (2004) recommended the use of the autoregressive integrated moving averages (ARIMA) procedure originally devised by Box and Jenkins (1976) or AUTO REG in some sources instead of ITSA or ITSACORR as an alternative method of analyzing single-subject time-series designs.

Although examples of the use of the ARIMA procedure were not found in the Speech-Language Pathology literature, it has been successfully used for decades in fields such as medicine, management, agriculture and economics to estimate effectiveness and to forecast trends (Wilson & Keating, 2007). A number of distinct advantages render it ideal for single-subject, time-series data utilized in single-subject design. As can be seen in Figure2., the ARIMA procedure (Box and Jenkins, 1976) includes all data points, comparing each performance value to the previous performance value thus encompassing the entire data set rather than only the pre-post treatment data as utilized in Cohen's *d* statistical analysis. That said, the applicability (and accuracy) of the ARIMA procedure depends on the mapability of the data to a model in the ARIMA method. This critical mapping step involves three basic phases: 1) identification of the model 2) estimation and correction of the model and 3) diagnostic checking and application of the optimal ARIMA model to forecast a time series.

A number of recently developed software programs help to perform ARIMAs. SAS 9.1™ (SAS Institute, Inc., 2007), SPSS 16.0 (SPSS, Inc., 2007) and ForecastX™ (Galt, 2003*)).* Among these, ForecastX™ is the most user- friendly, particularly to those who are less proficient in SAS or SPSS. Originally designed for students in econometrics and forecasting, Forecast X™ works in tandem with Microsoft Word Excel™. Based upon the recommendations of Box and Jenkins, model-mapping using the ForecastX™ program consists of the following steps to select between an AR (1,2,3) an MA (1,2,3) or an ARIMA (1,2,3) model. Figure 3 offers a schemata for the Box-Jenkins method for ARIMA.

1.) Model identification: the data is entered into Excel and the ForecastX™ program.
2.) Using the ForecastX™ "Analyze" function will provide correlograms for both auto-correlation function (ACF) and partial auto-correlation function (PACF). The ACF and PACF can be used to allow the researcher to see trend as seen in Figures 4 and 5.
3.) Estimation and forecasting: in this step the model hypothesized to be the best fit is analyzed by the program. The ForecastX™ will provide both Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values which can be utilized to adjust the model and arrive at the best model for the data as illustrated in Figure 5.

4.) Forecasting: Forecast$^{X™}$ will forecast the time series to a specified number of data points. Of particular interest to this application are a) the r^2 value or the fraction of variance explained by the model.

To illustrate the value of the methodology, the ARIMA procedure was utilized for analysis of a group of 20 single-subject studies of adult apraxia of speech. Results indicated an *r*-square of .923 a *p*- value of <.0001. These findings support the results of Strom and Boutsen (2006, 2007). They reported an *r*-square value of .745 and a *p*-value of <.0001 using Cohen's-*d* methodology . In their study, Strom and Boutsen raise the question of probable publication bias among the studies chosen. While this remains possible, the replication via the ARIMA procedure lends support to the conclusions supporting the current treatment efficacy in AOS. Another application of the ARIMA procedure is forecasting of future data patterns. In the time-series analysis the forecasted progress, (or in this case the results had treatment been continued), can be compared with the maintenance data provided representing the retention of skills after treatment was stopped. The data was analyzed in SAS and it was found that 13/20 data sets resulted in a forecasting value which was greater than the maintenance data, the difference was statistically significant (t=3.1142; p=.009). In treatment evaluation, even the results indicating that the data did not show forecasted improvement with additional treatment is useful to the clinician. It can be interpreted as an indication that perhaps progress had plateaued and additional treatment would not be efficacious. Such information could prove to be valuable to clinicians, third-party payers, and researchers in evaluating the efficacy of a treatment approach.

The ARIMA procedure is not without challenges to the aphasiologist and researcher. A possible challenge concerns the number of data points available in many single-subject studies. The Box-Jenkins ARIMA procedures are better suited to longer range rather than shorter range forecasting and larger rather than smaller data plots. However, Wilson & Keating (2007) stated that in actuality, ARIMAs have been used successfully for large, medium, and smaller data sets. For example, the ForecastX™ program requires at least 10 data points to function but has no upper limit.

Regardless of this license, ARIMAS are thought to work best when used on large data sets with lower volatility. Another limitation for clinicians and researchers in the area or speech/language pathology might be the advanced statistical knowledge required to reliably determine the correct model for ARIMA. However, the advent of user-friendly computer software such as Forecast X® (Galt, 2003), SAS 9.1, and SPSS 16.0 has simplified the process allowing students, and researchers new to ARIMA to reliably perform the statistical computations and arrive at an accurate conclusion.

With the consideration of these limitations, it seems feasible to use ARIMA procedures with the time-series data in a single-subject multiple-baseline design in the field of speech/language pathology.

Figure 1. The formula for determining pre-post treatment effects using Cohen's d statistic.

Cohen's *d* – *(Dunst, Hamby & Trivette, 2004)*

$$d = \frac{(\mu Maintenance - \mu Baseline)}{SD_P \,/\, \sqrt{2(1-r)}}$$

Where $SD_P = \sqrt{\underline{SDB^2 + SDM^2}}$

Figure 2. Comparison of Cohen's *d* and ARIMA values.

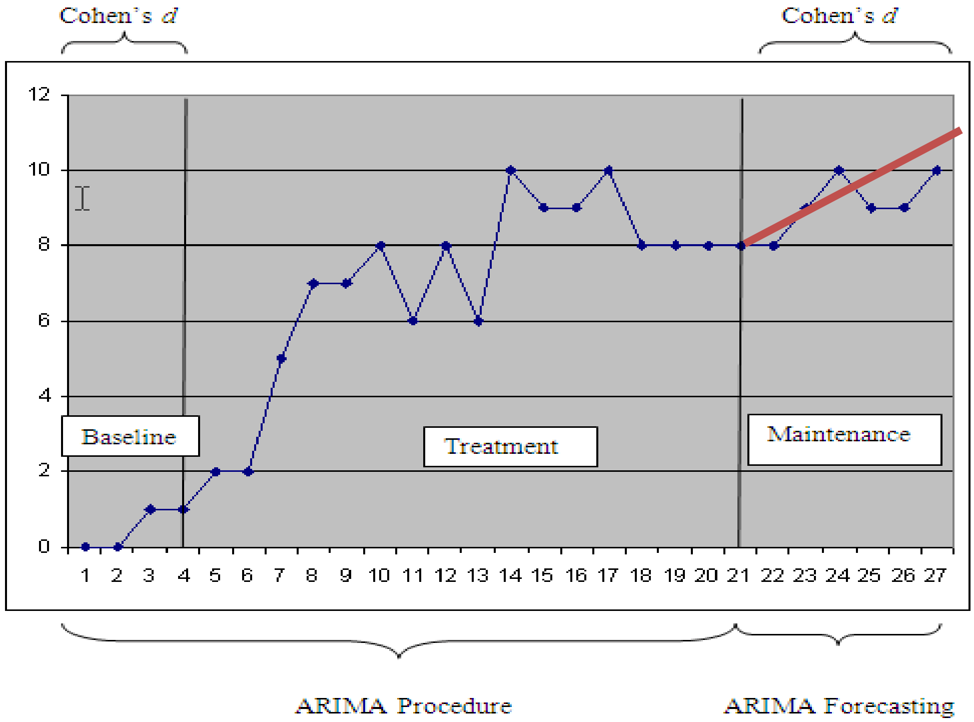
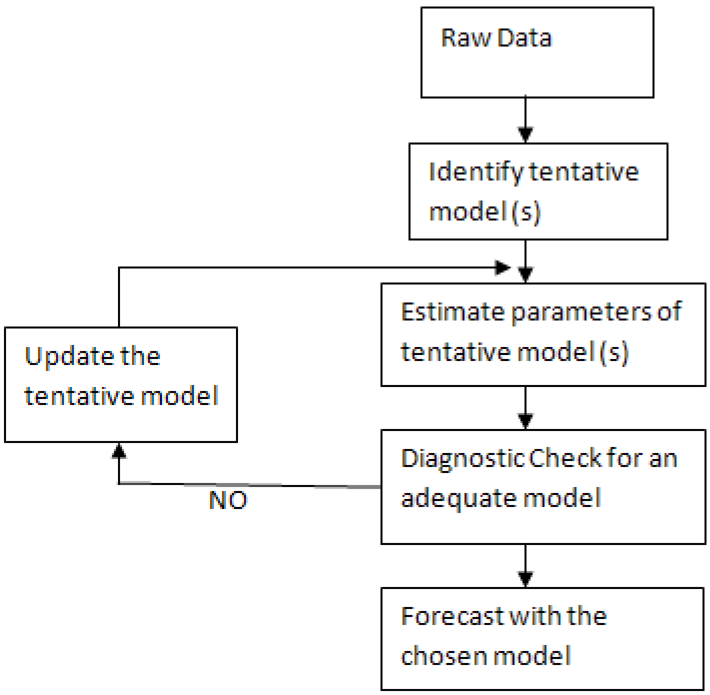

Cohen's d
Cohen's d
12
10
8
6
4
2
0
Baseline
Treatment
Maintenance
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27
ARIMA Procedure
ARIMA Forecasting

Figure 3.   A schemata of the procedure for ARIMA model building



.

Figure 4.   An example of ForecastX ACF/PACF values

ACF @ PACF - 3192008 Values For Series 1

Obs	ACF	PACF
1	.7352	.7352
2	.5887	.1047
3	.3462	-.2604
4	.1153	-.2197
5	.0406	.1929
6	-.0158	.0902
7	-.0434	-.1185
8	-.1050	-.2360
9	-.2125	-.1796
10	-.2675	.0804
11	-.4013	-.2037
12	-.3435	.1251

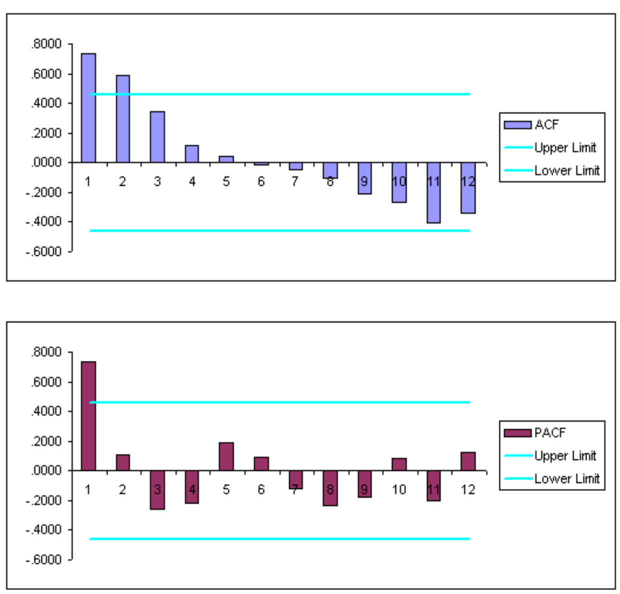


Figure 5. Auto-Correlation Function (ACF) and Partial Auto-Correlation Function (PACF) models (Wilson & Keeting, 2004)

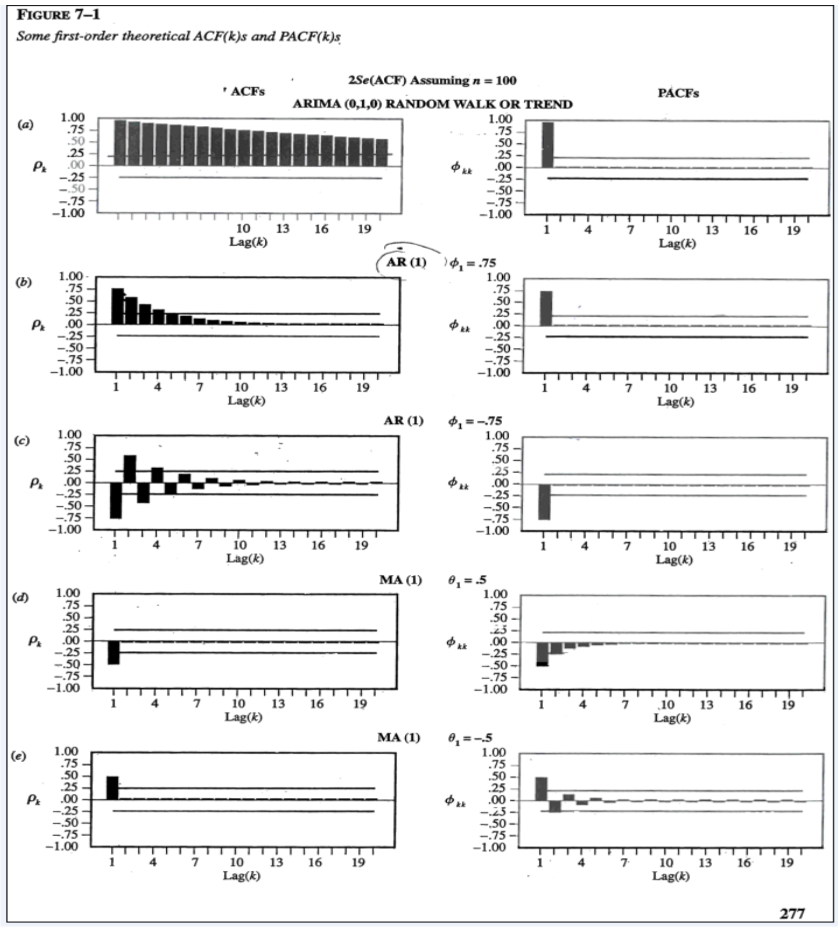


Figure 6. AIC and BIC Values.

ARIMA Model Forecasted (Box – Jenkins Method)	AIC Value	BIC Value
Autoregressive integrated moving averages or ARIMA (1)	240.99	243.58
Auto-regressive moving averages or ARMA (1)	241.94	245.83
Auto-regressive or AR (1)	241.94	245.83
Moving averages or MA(1)	241.94	245.83

Table 1. Comparison of ARIMA average forecasting data and average maintenance data provided.

Author	ARIMA Forecast Average	Maintenance Data Average	Greater Value
Kearns (1986)	7.768	7.4	ARIMA
Lustig & Thompkins (2002)	75.06	89	Maintenance
Cherney, et al. (1983) - set 1	15.416	12.66	ARIMA
Cherney, et al. (1983) - set 2	19.52	17.5	ARIMA
Raymer & Thompson (1991) - set 1	6.356	8.5	Maintenance
Raymer & Thompson (1991) - set 4	9.923	6.9	ARIMA
Simmons (1978) - set 1	87.28	77.5	ARIMA
Wambaugh, et al. (1996) - set 2	70.76	80	Maintenance
Wambaugh, et al. (1996) - set 3	80.57	70	ARIMA
Wambaugh, et al., (1998) - set 1	98.738	94.5	ARIMA
Wambaugh, et al. (1998) - set 2	86.42	94.75	ARIMA
Wambaugh, et al. (1998) - set 3	90.65	91	Maintenance
Wambaugh, et al. (1999) - set 1	93.653	65.5	ARIMA
Wambaugh, et al. (1999) - set 2	63.834	47	ARIMA
Wambaugh, et al. (1999) - set 3	13.325	24	Maintenance
Wambaugh & Martinez (2000) -set 1	92.66	92	ARIMA

Literature cited:

Box, G.E.P. & Jenkins, G.M. (1976). *Time-series analysis: Forecasting and control (2nd Ed.)* San Francisco: Holden Day.

Crosbie, J. (1989). The inappropriateness of the C statistic for assessing stability of treatment effects with single-subject data. *Behavioral Assessment, 11*, 315-325.

Dunst, C.J., Hamby, D.W., & Trivette, C.M. (2004). Guidelines for calculating effect sizes for practice-based research syntheses. *Centerscope: Evidence-Based Approaches to Early Childhood Development, 3*, 1-10.

Galt, J. (2003). *ForecastX Wizard* (Computer software. Chicago, IL: John Galt Solutions, inc.

Huitema, B.E., & McKean, J.W. (2000). A simple and powerful test for autocorrelated errors in OLS intervention models. *Psychological Reports, 87*, 3-20.

Huitema, B.E. (2004). Analysis of interrupted time-series experiments using ITSE: A critique. *Understanding Statistics, 3*, 27-46.

Strom, K.E., & Boutsen, F.B. (2006, November). *Treatment of acquired apraxia of speech: A qualitative/quantitative review.* Paper presented to the annual meeting of the American Speech-Language-Hearing Association. Miami Beach, FL.

Strom, K.E., Boutsen, F.B., & Christman, S.E. (2007, May). Treatment of acquired apraxia of speech: A qualitative/quantitative review of single-subject studies. Poster session presented at the Clinical Aphasiology Conference, Scottsdale, AZ.

Tryon, W.W. (1982). A simplified time-series analysis for evaluating treatment intervention. *Journal of Applied Behavior Analysis, 15,* 423-429.

Wilson, J.H. & Keating, B., (2007). *Business Forecasting* (4th Ed.). Boston: Irwin, McGraw-Hill.