Social Influences on Female Speakers' Pitch

by

Julie Anne Lewis

B.A. (Pomona College) 1994
M.A. (University of California, Berkeley) 1998

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy
in

Linguistics

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:

Professor John J. Ohala, Chair
Professor Robin T. Lakoff
Professor Yoko Hasegawa

Spring 2002

Social Influences on Female Speakers' Pitch

by

Julie Anne Lewis

Abstract

Social Influences on Female Speakers' Pitch

by

Julie Anne Lewis

Doctor of Philosophy in Linguistics

University of California, Berkeley

Professor John J. Ohala, Chair

This dissertation discusses two experiments isolating interlocutor influences on females' pitch in conversations. They explore the effects of interlocutor gender and status and conversation topic. Subjects were twelve Caucasian, heterosexual, female college students, speakers of California English. Each spoke with four unfamiliar interlocutors, a female and male peer and professor. In experiment one, solid-minute excerpts from the start of the interview and from two kinds of topic were created for each interview. Median pitch, standard deviation of pitch and 80% pitch range were calculated. In the pooled data, pitch range was significantly larger with female interlocutors compared with males ($F(1, 73) = 6.179$, $p < 0.05$). Data from several individual subjects echoed this finding or showed larger standard deviations of pitch (both indicating more varied pitches) with females, and two had higher medians with females. These effects may reflect a camaraderie/high engagement politeness style used between American females. Regarding interlocutor status, two subjects had higher median pitches with professors, possibly reflecting deference politeness, and two others had more varied pitches with peers, showing camaraderie with them. Only minimal effects due to topic were found.

Experiment two explores whether interlocutors affect speakers' pitches via accommodation, measuring whether two people's pitches correlate within conversations. The level of correlation was measured for all interviews for all three pitch measures in two different ways, using normalized scores based on averages within separate minutes of the interviews, and using a delta measure based on differences between scores from consecutive minutes. It was found that, for the most part, accommodation cannot account for the patterns found in experiment one. However, when significant correlations in individual interviews are grouped together, some patterns arise. Many occurred for median pitch, and more were based on delta measures with female interlocutors while more were based on normalized scores with males. Pitch accommodation is interpreted as reflecting solidarity/camaraderie politeness. The predominance of positive correlations (only a few negative ones were found) in experiment two, in tandem with the heightened variability with females found in experiment one, was taken as evidence of the importance of positive politeness strategies among these female subjects.

Table of Contents

Acknowledgements

I could not have completed this dissertation without the help of many people. My parents, Margery and Larry Lewis, have always encouraged me to follow my dreams and have supported me in countless ways. My siblings Laura, Steve, and Dan, have cheered me on all these years, and my twin sister Janet has kept track of my life and work and been my faithful telephone companion and best friend through it all. Long-distance emotional support also came from Lani Golay, Gina Moffitt, and Stephanie Vernooy.

I thank my committee, John Ohala, Robin Lakoff, and Yoko Hasegawa, for their support and feedback. My advisor John Ohala taught me about phonetics and about being an academic and a good scientist, and inspired me with his creative thinking. He also taught me that methodologies should be driven by the questions being asked, rather than by the methodologies that are used standardly or that the scientist is comfortable with. Robin Lakoff was the reason I chose UC Berkeley; her work has shown me how relevant linguistics is to our lives as humans. Her approval and encouragement have been especially gratifying for this reason. I would also like to thank Leanne Hinton and John McWhorter for encouraging me to pursue this topic at the sociolinguistics lunch meetings a few years back.

Other colleagues have also helped me in crucial ways. I could not have emerged from my statistical analysis in one piece without the help of Lily Liaw, who graciously and patiently explained SPSS to me. Our discussions about linguistics, feminism, and countless topics have enriched my life and my dissertation. I would also like to thank the women who went before me in the Phonology Lab for showing me the way, especially Natasha Warner and Madelaine Plauché. Suzanne Wertheim, my dissertation-writing accountability partner, has listened to my struggles and offered advice, encouragement, and friendship throughout, but especially during this last year of analysis and writing. Other linguists who have helped with useful discussions, advice, and support over the years are: Ikuko Yuasa, Maria-Josep Solé, Ian Maddieson, Emily Bender, Andy Dolbey, Gunnar Hansson, Ben Bergen, Josh Guenter, Margaret Urban, Aimée Lahaussois Bartosik, Ashlee Bailey, and Mo Corston-Oliver.

I will be forever grateful to Ronald Sprouse for writing perl and Praat scripts and for technical advice throughout, as well as to Jos Pacilly of Universiteit Leiden for sharing his scripts that inspired Ron's. Thanks also to the UCB statistics graduate students involved in the statistical consulting service for providing very useful advice.

John Ohala, Pam Morgan, Andrea Frome, and Ben Bergen deserve special thanks for their fine work as my experimental conspirators and interviewers. I would also like to thank the subjects who participated in the study.

For financial support, I am grateful to the UC Berkeley Linguistics Department, which supported me in part during my first year and during the semester in which I advanced to candidacy, and to the P.E.O. Sisterhood (through the recommendation of my aunt, Mary Helfrey), which supported me during a semester of dissertation research. The staff of the Linguistics Department, Belén Flores, Esther Weiss, and Paula Floro, have also always been extremely helpful in keeping the logistics of my graduate school career on track. Whenever I asked Belén for something, which was often, she was always helpful and always in a good mood.

Lastly, but not in any way in the least, I would like to thank my partner Alex Papavasiliou. He has supported me emotionally and intellectually during these last

harrowing years of graduate school, listening to my experimental and analytical woes *ad nauseum* and being in general a wonderful person. I dedicate this dissertation to him.

Chapter 1. Introduction: On Explaining the Pitch[1] of Women's Voices

1.1. Gender Differences in Vocal Pitch

Average speaking fundamental frequency (SFF) has consistently been found to vary between men and women. The finding that men's average SFF is lower is extremely robust, but different studies have had slightly different findings regarding the extent of this difference. Künzel (1989) measured read speech in 105 men and found F0 averages and medians of 115.8 Hz and 113 Hz; for 78 women, they were 210.6 Hz and 210 Hz. Hollien and Jackson's (1973) data for 157 American young (17-25 years old) males were 129.4 Hz for read speech and 123.3 Hz for extemporaneous speaking. Shevchenko's (1999) data for Americans included women's mean F0s: 185 Hz for upper class, 227 Hz for middle class, and 198 Hz for lower; men's were 102 Hz, 111 Hz, and 126 Hz, respectively. Eklund and Traunmüller (1997) found that mean F0s were 109.4 Hz for men and 206.2 Hz for women.[2] Thus, women's average SFF is significantly higher than men's, but the above studies measure the difference between the sexes as being anywhere from about 85 to 115 Hz.

Other pitch variables besides average SFF have been studied with an aim to clarifying the differences between men's and women's use of pitch. Pitch range, dynamism (rate of pitch change), and pitch variability (standard deviation of average SFF) have all been studied, but the results of these studies are more controversial. Impressionistically, many

---

[1] Pitch is a perceptual phenomenon, but is related to fundamental frequency, an acoustic and more easily measured phenomenon typically measured in Hertz (cycles per second). I will use pitch as a cover term throughout and specify units of measurement.

[2] There were pitch differences between the different vowels, i.e., intrinsic pitch. Men's average /ɒ/ F0 was 101 and women's was 188, but for /i/, men's was 129 Hz and women's was 232 Hz. When converted to semitones, however, the intrinsic pitch effects looked similar for men and women.

have noted that women tend to use a wider pitch range than men do. However, when Henton (1989) explored pitch range in semitones (a logarithmic scale more reflective of how we perceive pitch) instead of Hertz as previous studies had, she found no significant differences in men's and women's pitch ranges in read speech for American speakers. She asserts that expectations and stereotypes about women's speech do not reflect their actual behavior. However, Shevchenko (1999) found that women's ranges in semitones were 8, 7, and 6 for upper, middle and lower classes, but men's were 6 for all three classes. Cross–cultural differences are a possibility, but Yuasa (2000) studied perception of pitch ranges using the ERB (Equivalent Rectangular Bandwidth, a partially linear and partially logarithmic and even more perceptually accurate) scale and found that pitch ranges of men and women are similar in Japanese (personal communication). Henton (1995) also studied pitch dynamism (using semitones) to see how quickly changes occur within the pitch range used. She found no significant differences between men and women, and her data for English and French were similar. However, Haan and van Heuven (1999) did find some pitch range differences between male and female Dutch speakers. After hypothesizing that the lack of gender-based pitch range differences in studies' findings may be due to the studies' focus on declarative sentences, they found that their female speakers had significantly wider pitch ranges locally and globally in read questions *and* statements than their male speakers did (five male and five female speakers). Their preliminary study of spontaneous speech replicated the findings for questions.

In addition to discussing the controversy over pitch range, Henton (1995) also mentions that the separate factor of pitch variability differs in men and women. She cites

Tielen (1992), who found that women's pitch variability, measured via standard deviation of F0 from linear regression lines drawn through pitch contours of intonation units in read sentences, was on average 0.3 semitone larger than men's for 10 female and 10 male Dutch speakers. This difference was not statistically significant, however. Künzel (1989) measured the standard deviation around the average F0 for 105 men as 16.5 Hz and as 17.3 Hz for 78 women. Furthermore, regardless of these types of statistical pitch measures, gender-based intonation differences are entirely possible, as suggested by Brend (1975), who, using intonation patterns reported in Pike (1945) but not presenting acoustic data, claims that women have more levels of contrast in their intonation contours, and McConnell-Ginet (1978), who found more rising intonation patterns in women.

Thus, there are indeed some gender differences in pitch, although scholars debate over the nature and extent of them. What remains to be answered is what are the causes of these differences. Gender and its concomitants are often described categorically but not explained. Smith (1985) discusses the idea that humans have a proclivity for categorization, in order to make a complex environment easier to deal with. Categorization helps in predicting behaviors, but it can lead to insensitivity to variation within categories and an emphasis on between-category differences. He describes "psychological hypostatization" in which "the categories in the eye of the beholder are endowed with an independent ontological status, an apparent intrinsic reality, apart from the processes that lead to their emergence and recognition in society" (p. 20). If the categories (of gender) are self-evident, they need description, and not explanation—Smith sees this as a problem. In the past, many studies have merely described differences between male and female speech and not questioned their origins. The important task remains to ask what is behind these differences—to explain them, and not just describe them.

In the following sections, I will discuss the previous work on explaining gender differences in pitch, and how they led to the design of my study. The two main candidate arenas for explaining gender-based pitch differences are physiology/biology and social/societal influences. Studies will be grouped according to whether their hypotheses pertain to biological or social explanations.

1.2. Biological Differences

Sexual dimorphism can be found in the vocal apparatus of adult males and females. During puberty, the cartilages of the larynx grow rapidly in males and their vocal folds increase by about 10 mm in length and thicken, while females' folds only increase by about 4 mm, and their larynxes grow at the same rate throughout childhood and puberty. Thus, male postpubertal vocal folds are between 17 and 20 mm in length while female folds are between 12.5 and 17 mm (Zemlin, 1988, p. 176). Zemlin reports the findings of several studies that compared the angle of the thyroid laminae of the larynx in males and females that all found males to have a smaller angle though the size of the difference varied considerably in the different studies. (The thyroid cartilage provides the anterior attachment for the vocal folds.) This smaller angle creates a more prominent thyroid eminence (Adam's apple) in males. However, one study (Smith, 1978, reported in Zemlin, 1988) found that there was not a consistent relationship between vocal fold length and the thyroid angle. However, regardless of how this vocal cord length difference occurs, its existence is uncontroversial.

The larynx is also lower in the vocal tract in adult males than in adult females. Throughout childhood, the larynx descends within the vocal tract in boys and girls due to growth of the vertebral column and changes in the angle between the skull base and the vertebral column (ibid., p. 175). However, at puberty, the growth patterns diverge; the distance between the roof of the nasal cavity and the hyoid bone, which the larynx is connected to, increases rapidly in males but not in females (Goldstein, 1980, cited in Ohala, 1994). The result is that the adult male vocal tract is 15-20% longer.

These sexually dimorphic characteristics of the larynx and vocal tract lead to certain phonetic differences. Since longer, more massive objects have lower natural frequencies, one would expect males to have lower fundamental frequencies and lower formant frequencies. Indeed they do. However, the issue of the extent to which the above anatomical differences are responsible for the phonetic differences in men's and women's voices is far from completely understood. There has not been extensive research clarifying the relationship between dimorphic vocal apparatus and speech differences between men and women, and the findings of this research have been somewhat contradictory. Methodological issues have forced researchers to make inferences that make their results somewhat difficult to interpret. For instance, it is difficult to obtain accurate measurements of the dimensions of the vocal cords (and vocal tract) in living subjects without using invasive techniques. Thus, researchers have attempted to correlate fundamental frequency with vocal cord length in various indirect manners. Again intuitively, it makes sense that various bodily dimensions would positively correlate, and thus many have assumed that a person's overall size, especially his or her height, would be a good predictor of larynx size and thus vocal cord length: "A tall well-built man will tend to have a long vocal tract and large vocal folds. His voice quality will reflect the length of his vocal tract by having correspondingly low ranges of formant frequencies, and his voice dynamic features will indicate the dimensions and mass of his vocal folds by a correspondingly low frequency" (Laver and Trudgill, 1979, pp. 7-8, cited in Graddol and Swann, 1989, p. 16). Furthermore, listeners appear to be able to judge a speaker's overall physical build based on voice samples with moderate success. Fay and Middleton (1940) found that their subjects could rate a speaker's Kretschmerian constitutional type (body type) after hearing a read passage at a level higher than chance. They hypothesize that certain voices reflect stereotypes and are thus judged more consistently than others; voices of pyknic (glossed as "fat, portly") types were judged at a level 22% above chance and voices of leptosomatic (glossed as "slender, skinny") types were judged at a level 20% above chance, but athletic types were judged only at a level 1% above chance. It is interesting that listeners were fairly accurate—it could be that the stereotypes they hold are not completely unfounded—but there were only nine speakers, three of each type (judged by the authors) in this study. It is also unclear what cues the listeners were reacting to. However, in a study with 28 speakers, Lass et al. (1980) provide some evidence regarding what cues listeners are using in a similar task. They found that listeners were indeed able to identify speakers' height and weight fairly accurately based solely on recorded read speech samples, but their identification was not significantly less accurate when they were presented with whispered speech (although inter-rater reliability did suffer slightly for the whispered samples). If whispered speech can be used to predict these physical characteristics, too, then vocal pitch cannot be the only or even necessarily the crucial cue for listeners.

Studies that have tried to substantiate claims about the connection between speakers' physical characteristics and vocal pitch more directly have not been consistently successful, either. Künzel (1989) obtained height and weight data (via self report) and measured average fundamental frequency based on read passages for 105 male and 78 female adult subjects. He found no correlation between F0 and height or weight in men or women, and concluded that somatic information is not carried in the speech signal. (Height and weight did correlate positively for both men and women, so at least in this way, one physical dimension correlated with another in the expected manner.) Graddol and Swann (1983) had slightly different results. Since their critical review of the literature suggested that there is not convincing evidence for body build being associated with pitch level (or for the claim that listeners can judge speakers' height and weight from vocal cues), they sought to put the issue to rest by taking their criticisms of previous work into account, controlling for social variables and using longer speech samples. They recorded 12 males and 15 females (socially homogenous) reading two different passages, one neutral in content and one a dialogue requiring more inflection. They also recorded subjects phonating the vowel /a:/ at their lowest possible pitch. They then weighed and measured them and calculated mean F0, median F0, and basal F0 for each speaker. They found no correlation between any pitch or size measurement for the females, but for the males, there were two significant correlations: height correlated (negatively) with median F0 in the neutral passage and with basal F0. However, these results can be interpreted in several ways. One interpretation is that the men were using an F0 that was reflective of their larynx size, i.e., a "natural" F0 (p. 358). Given that basal F0 did not correlate with height for the females but did for males, two separate findings are suggested, a) median F0 is related to physical factors for men but not women, and b) that height does not reflect larynx size (or at least the physical mechanisms that affect F0) in women but does in men (p. 360). One possible flaw in this interpretation is that it assumes that the basal F0 as measured reflects larynx size. But, it may be that it was not reliably measured; some subjects may be better at achieving their actual lowest possible pitch than others. The authors themselves discuss another possible problem that could lead to an alternative explanation. The two kinds of reading passages evoked different intonation patterns in that the first was of a tedious nature that "did not encourage speakers to use large inflections of frequency" and had long declination patterns as lengthy and complex sentences tend to have (p. 361). The correlations between F0 and height varied depending on the intonational contour of the sentence, with the declining parts of the passage having stronger correlations. Thus, the findings may have been influenced by women using more variable intonation patterns and (thus less declination). An explanation could be that the males were not using a more "natural" F0, but that they were using the lower portions of their possible ranges and using a narrow range for social and/or ethological reasons (e.g., to seem larger by maximizing their apparent height; see Ohala, 1994, for an ethological explanation of pitch usage). Nevertheless these findings are intriguing.

The above studies attempted to study the extent to which anatomical constraints determine F0 indirectly, and thus their inconclusive (and even negative) findings may not tell the full story. Hollien and Jackson (1973) have more direct evidence in that they obtained measurements of the anterior-posterior dimension of the larynx via lateral x-ray, as well as height, weight and other body size measurements, in 157 adult males. They measured average F0 in read and extemporaneous speech and possible phonation range. Surprisingly, they found no statistically significant correlation between any voice parameter measured and any size measurement, including larynx size. Not even the two correlations found in Graddol and Swann's (1983) data for males were found. Two possible confounders are the topics of the extemporaneous speech samples (subjects gave a three-minute talk on one of four different topics, and the different topics could have evoked different intonation patterns) and the subjects' being told that it was their voices being studied in order to help them feel less self-conscious about their speech content, which could have led them to be self-conscious about their speaking voices. Thus, there appears to be little to no *direct* positive evidence for individuals' average fundamental frequencies correlating with their physical sizes, even their larynx sizes.

However, Ingo Titze's (1989) rigorous review of physiologic and acoustic differences between males and females points to possible explanations for the negative results of the above studies. In particular, he points out that larynx length is not the same as vocal cord length. Citing data from Kahane (1978), he asserts that the thyroid cartilage is 20% larger in males than in females in the anterior-posterior dimension, but the membranous vocal folds are 60% longer in males. The 20% difference is found in other laryngeal dimensions (lateral, vertical), which suggests that the membranous portion grows "disproportionately" (p. 1700). The length of the larynx (thyroid cartilage) includes that of the vocal fold processes as well as the muscular portion behind the glottis. Titze maintains "[t]his disproportionate growth, together with the unknown growth rate in the muscular portion behind the glottis, makes an inference about L_m [membranous length] on the basis of L_l [overall larynx length] somewhat uncertain" (ibid.). It may be the case that Hollien and Jackson found no correlation between larynx size and fundamental frequency because they were measuring overall larynx length. In fact, Titze found that the biggest scaling factor in explaining mean F0 differences between males and females is that based on membranous vocal fold length. (This factor also accounts for mean airflow and aerodynamic power differences.) Vocal cord thickness, on the other hand, is much less important in the manipulation of fundamental frequency because as the mass increases, stiffness does as well, and these two factors counterbalance each other.

Kaneko et al. (1983) found that there are some sex differences in the resonance characteristics of the vocal folds themselves, and that the folds behave differently when they are at rest versus when they are engaged and tense. During quiet respiration, in normal male adults, the resonant frequencies of the vocal fold (measured via response to a vibrator system) during quiet respiration ranged from 91 to 145 Hz (average 128 Hz), and in normal females, they ranged from 115 to 167 Hz (average of 136 Hz); females' vocal folds tended to have slightly (but only slightly) higher resonant frequencies, perhaps reflective of their smaller mass. However, during actual phonation, "the biomechanical factors of the vocal folds have to change depending on muscle control" (p. 304), so in order to better understand differences in the mechanism of phonation, measurements were made of folds in phonation neutral position (speakers sustaining the vowel [i] stopped voicing but maintained the same laryngeal gesture just before measurements were made) at various pitches (p. 316). In the examples discussed, both males' and females' vocal folds had two resonances, one at about 100 Hz and the other somewhat near the phonation pitch, which was 150 Hz for a male subject and over 200 Hz for two female subjects. The authors explain their findings: "During quiet respiration, the vocal fold is relatively relaxed. However, once the vocal fold is tensed, such as in the phonation neutral position, the mechanical properties are changed significantly, sex being a factor" (p. 307). Once the folds are tensed for phonation, they behave differently in males and females, leading to different phonation pitches and resonant frequencies of the folds themselves.

The above discussion on vocal fold/larynx size and its capacity for explaining pitch differences between men and women does not tell the full story. It is difficult to tease apart the effects of fundamental frequency and the resonances of the vocal tract—both contribute to the perception of the pitch of a person's voice (cf. Coleman, 1971, 1976), which is why when a person breathes in helium and speaks, his/her voice sounds as if it were at a higher pitch (vis-à-vis when breathing air) even when he/she is phonating at the same fundamental frequency. The longer vocal tract in males mentioned above has phonetic ramifications, leading to lower resonant frequencies of the vocal tract, i.e., men have lower formant frequencies. Fant (1966) provides acoustic and physical data that elaborate on the nature of the 15-20% difference between the sexes in the length of the vocal tract. He asserts that acoustic analysis of vowels in Swedish and American English demonstrates that female formant frequencies are not related to those of males by a simple scale factor reflecting the overall vocal tract length. Different vowels have greater and lesser sex-based differences, namely, F1 and F2 of back rounded vowels (and F1 of high front vowels) have a smaller scale factor while F1 of low vowels diverges more than average. Fant claims that "[t]he main physiological determinant of the specific deviations from the average rule [simple scale factor rule] is that the ratio of pharynx length to mouth cavity length is greater for males than for females and that the laryngeal cavities are more developed in males" (p. 22). Using acoustic measurements of vowels, Fant inferred cavity lengths that matched actual physical measurements (Fant's measurements from P. Edholm's unpublished x-ray data and measurements from Chiba and Kajiyama, 1941) fairly well. Thus, the fact that the ratio of male pharynx length to female pharynx length is approximately 1.3 while the ratio of male mouth cavity length to female mouth cavity length is approximately 1.18 appears to explain the vowel category-dependent nature of formant differences between men and women. The proportionally larger pharynx length in adult males follows from the sexually dimorphic lowering of the larynx in males at puberty. These non-uniform (and anatomically-based) formant differences "may not have a very crucial importance for the phonemic identity of perceived vowels in connected speech but are undoubtedly of interest as speaker category determinants" (p. 35).

The above discussion has described the avenues explored and techniques used in exploring the relationship between physical parameters of men's and women's larynxes and vocal tracts and their F0s and formant frequencies. While some progress has been made and answers suggested, it is apparent that methodological difficulties hinder efforts to find definitive answers as to the role of anatomy and physiology in explaining gender differences in pitch. Other studies have attempted to find the explanations for these differences by isolating and exploring social influences on pitch.

1.3. Social Influences: Group Membership

1.3.1. Children's Speech

An indirect way of exploring the idea that anatomical differences are not solely responsible for the phonetic differences between men and women is to study children's speech, as "[t]here are no discernible sex differences in the infant and children's larynx" (Zemlin, 1988, pp. 175-6), so if there are differences between girls' and boys' speech, it must be socially motivated, if they indeed have similar vocal anatomy. Fant's (1966) review of acoustic measurements of vowels in children also suggests that the pronounced (dimorphic) larynx lowering happens only at puberty (in males). He asserts that children's formant patterns are scaled proportionally relative to female's formant patterns, suggesting that children and women share pharynx length to mouth cavity length ratios[3]. Sachs, Lieberman, and Erickson (1973) note that listeners can categorize voices as either male or female and question whether this ability is due to anatomical differences or culturally-determined factors (considering phonetic, not lexical differences). Citing Mattingly (1966), they assert that the acoustic differences in F0 and formant frequencies in adults are too large to be explained by the sexual dimorphism that arises at puberty and must be based in part on adaptation to male and female archetypes (that men are big and women are small). Since acculturation occurs in childhood, the study tests whether or not children's speech reflects gender differences. If height/weight differences between boys and girls are controlled for and they have similar mandible lengths (although that pertains to only half of the vocal tract, Fant's (1966) data suggests that there isn't sexual dimorphism in pharynxes in children), their larynxes and vocal

[3] The children's data that Fant (1966) discussed (from Peterson and Barney, 1952) is not presented separately for boys and girls.

tracts are probably similar, so their F0s and formants should be similar, if only anatomy is coming into play. Twenty-six children (4-14 years old, 14 boys and 12 girls) were recorded reading a sentence and a passage and sustaining the vowels /a, i, u/. Eighty-three adult listeners heard the randomized sentences and judged the sex of the speaker. Listeners could indeed distinguish the boys and girls—they were correct 81% of the time (significant). Acoustic analysis showed that the boys actually had significantly higher average F0s (274 Hz) than the girls did (249 Hz), but the first two formants were significantly lower for the boys, for /i/ and /u/, just as in adults:

	/i/		/u/		/a/	
	F1	F2	F1	F2	F1	F2
Girls	321	3247	420	1173	968	1568
Boys	302	3136	352	975	932	1611

Misidentified girls tended to have boy-like formants, although they weren't bigger. (They were also described as "athletic" and "tomboys.") The authors explain the results as being due to boys and girls using their articulatory mechanisms differently (e.g., lip rounding lengthens the vocal tract, smiling shortens it, and different tongue configurations can alter formants) to mimic culturally-determined formant patterns of what is appropriate for the sexes. (For /a/, boys appear to be lowering formants by making a more central vowel. The authors assert that they are actively manipulating formants, as F1 and F2 aren't both consistently lower as boys get older/taller, which is evidence that the formants are not simply reflecting anatomical size [but this argument ignores the unclear relationship between height and vocal fold length].) Thus, apparently socially-motivated formant patterns seem to be at least in part responsible for listeners' ability to discern the gender of children by hearing their speech. However, formant frequencies certainly are not the sole explanations—listeners heard sentences with tempos, intonations, voice qualities, pitch ranges/variations, and pronunciations of words. Furthermore, it is interesting and a bit surprising that the boys' F0s weren't lower than the girls,' if the children were manipulating phonetic cues for social reasons. F0 is a somewhat manipulable cue. The data suggest boys and girls appear to try to sound masculine and feminine, but it is unclear how exactly they were doing this and what cues listeners were using when they correctly identified speakers' genders.

Given the findings of Sachs et al. (1973) that people could identify the sex of the speaker when hearing the voices of prepubescent children, Sachs (1975) explored what cues listeners might be using. Listeners didn't seem to have intuitions about how they knew and were surprised at how accurate they were. The previous study did not control for sentential cues in the voice samples, so a study was conducted using isolated vowels (the ones recorded in the 1973 study, with the children matched for height and weight). Listeners correctly identified the sex of the speaker 66% of the time, which is significantly above chance levels. When responses to individual voices were analyzed, it was found that listeners were using vocal tract size—the biggest children were thought to be boys and the smallest ones were thought to be girls. However, the children in the middle were judged correctly. Isolated vowels indeed carry some usable cues, but the listeners were less accurate than they were when they heard sentences, when they made judgments that were independent of age or size cues. To explore whether sentences were simply providing more exemplars or whether they were supplying other characteristics (rate, intonation), a second experiment was conducted playing the sentences backward. Listeners were correct 59% of the time (not significant). These results were not significantly different from the isolated vowel experiment, but they were significantly different from the results obtained with sentences played normally. It seems that sentential cues must be playing a significant role in helping listeners identify the sex of the speaker. However, the moderate success of listeners hearing isolated vowels does support the idea that some gender information is carried in (manipulated) formant frequencies.

Nairn (1995) found similar results with Scottish-English speaking children. In this study, 89 children (4.5-5.5 years old, younger than the children used in many previous studies) from two Edinburgh primary schools served as speaker-subjects. They were recorded saying isolated vowels (/i, a, o/, i.e., varying height and frontness), isolated sentences, and spontaneous passages controlled for topic (narrating/retelling a story using provided pictures). Listeners were eight males and eight females between 18 and 33 years of age. They were asked to rate the samples as male or female. Both the sentence and passage samples were significantly better identified than the vowel samples, but all identification rates were significantly above chance (sentences were identified correctly 76.23% of the time, passages approximately 73% of the time, and isolated vowels 65.91%). Female judges were better at identifying sex, and girls' voices were correctly identified more often than boys'. The correct identification rates are similar to those in other studies (66-74% on average). The author explains cross-linguistic differences in identification rates as being due to different (learned) methods of conveying gender. It is interesting that the (perceptually-usable) differences occur in very young children in this study.

Moore (1995) demonstrates that some cues indeed are language-specific. This study follows in part from Karlsson and Rothenberg's (1992) study that found that Finnish children's gender was harder to discern from voice samples than Swedish or English-speaking children's gender was. It aimed to test the accuracy of Finnish and American speakers in identifying the sex of 4-6-year-old Finnish speakers and to explore any prosodic differences. Recordings were made of pre-school Finnish children saying utterances that were two to eight syllables long and utterances of eight syllables or longer. Five girls' and five boys' recordings were used both in test 1, with the shorter utterances, and in test 2, with the longer utterances. There were three groups of listeners, 18 Finnish and 13 American adults, 25 Finnish kindergarten teachers, and 55 Helsinki grade school children (9-12 years old). All groups had trouble distinguishing the boys and girls. After only being correct 41% of the time in test 1, the Finnish listeners improved slightly to 49% in test 2 (the longer utterances). The kindergarten teachers were included to see if familiarity/experience with kids' voices mattered—they did poorly, too, but their accuracy improved significantly in test 2 (to 59% from 32% in test 1). The improvement in test 2 suggests that "Finnish intonational or prosodic patterns that are associated with gender require longer utterances than eight syllables for correct identification" (p. 300). However, overall, the study shows that pre-pubescent Finnish children's voices don't really have prosodic cues to gender, corroborating Karlsson and Rothenberg's (1992) finding that Finnish children's voices carry less gender information than children's voices in other languages. Karlsson and Rothenberg suggest that grammatical classifications at the lexical level (less salient in Finnish) could affect gender conceptualization which could have ramifications in voice quality distinctions being (not) present. This theory has not been proven. Regardless of the grammatical structure of a language, its gender concepts could affect voice pitch and quality (cf. Ohara, 1999). However, it could be that

Finnish gender roles are less strongly marked, or it could be that Finnish gender roles are simply expressed in ways other than through voice pitch or quality. The author offers another speculative explanation, that there is a tendency among Finns to avoid interacting with strangers. Since the experiments were carried out by people who were strangers to the children, they could have been reacting to that and speaking with "stranger" prosody that could override any gender differences (cf. Yuasa, 1999). The children's voices were described by Finns as "shy," "cautious," and "scared." Even accepting the lack of significant findings at face value, the results of this study are not counter-evidence for the idea that social factors are in part responsible for phonetic differences between the genders; they show just how social the influences can be, since they are different in different cultures.

1.3.2. Cross-Cultural Studies

Cross-cultural studies also provide evidence for non-biological sources of phonetic differences between men and women. Loveday (1981) found evidence of cultural differences in pitch in Japanese and British English. This study focused on pitch range in Japanese and English politeness formulae. Ten subjects (two female and three male Japanese speakers and two female and three male English speakers, ages 23-46) were recorded reading a certain role in a dialogue about meeting someone and being invited for lunch later involving several politeness formulae, imagining that their interlocutor was a non-intimate acquaintance (always played by the experimenter). Pitch ranges used in the formulae as well as possible phonational ranges were obtained. The measurements showed that the phonational ranges are (reasonably) similar within each sex and across the two cultures, so any differences in pitch in politeness formulae cannot be explained by them. Loveday asserts that there are "clear sex-based intonational differences in the expression of politeness formulae for Japanese speakers that do not hold for English speakers" (p. 82). Japanese women use a very high range (e.g., 450 Hz), while Japanese men do not, while in English, both men and women use a fairly high—and at times similar—pitch range in polite expressions. Japanese men do not exceed 120 Hz, while English-speaking men do. A few British men reached frequencies well over 200 Hz in the politeness expressions.[4] The author interprets the results as showing that pitch level in Japanese serves to highlight sex differences more markedly than it does in English, at least in polite expressions (p. 85). Japanese expectations of social and sexual roles are more rigid than English norms. High pitch marks deference/politeness and in Japan, that is more (consciously) connected to femininity in Japanese. Japanese men also seem to signal masculinity via low pitch. A possible criticism of the study is that the number of subjects was quite small. There were also difficulties in getting accurate readings from the pitch meter (measurements were admittedly too low in frequency).

Not all cross-cultural studies have found the results expected by researchers who believe that pitch differences are tied to culture-specific gender construction. Van Bezooijen (1995)'s study confirmed the hypothesis that Japanese subjects have a stronger differentiation between their concepts of the ideal woman and the ideal man than Dutch subjects do and the hypothesis that high pitch is preferred for Japanese women more so

---

[4] Some of the results contradict Yuasa, 1998, who found that both Japanese men's and women's pitch ranges depended in the same way on the level of formality of the interaction/the familiarity of the interlocutor; they both used a narrow pitch range with unfamiliar members of their society and a wider range with familiar members (outgroup vs. ingroup), while in this study, Japanese women seem to use a wider pitch range and not just a higher overall pitch than the Japanese men. Perhaps the differences in this study have to do with the content being politeness formulae.

than for Dutch women (according to ratings of manipulated tokens). Despite these findings on cultural differences and previous pitch findings to the contrary, in the recorded (unmanipulated) data collected, the difference in average F0 between Japanese women and Dutch women was not significant (Dutch, 180 Hz, Japanese, 185 Hz). However, the sample was rather small--eight women from each culture, and the measures were from a very short (approximately 13 seconds) read text. It may also be the case that social roles may be changing in Japan; the women in the study were all highly educated.

Other studies have found some interesting phonetic differences even within one language or culture. Shevchenko (1999) found that pitch can vary across dialects and even socio-economic classes within a language in a study where several F0 parameters were correlated with various social group factors. Subjects were 125 British speakers who created 250 samples (each 2-3 minutes long) of read and spoken speech. Overall, the author concludes that an increase in F0-range (which involved a lower F0-mean and a higher F0-variability in this study) and a slower tempo symbolize south vs. north, urban vs. small towns, higher class, middle age (vs. young), men, reading (vs. speaking), formality, and authority (vs. non-authority or solidarity). Results were somewhat different in the study of 69 American speakers who created 138 samples (reading the same text as the British speakers and a speaking a spontaneous monologue), but the Americans and British had the following in common: Wider F0-range, greater F0-variability, lower F0-mean and slower tempo mark higher class, middle age, and reading. The social items grouped together in the results (including being male), if they have a unifying factor in common, all may involve some kind of higher social status. Furthermore, while British women had narrower pitch ranges (measured in semitones) than British men, the opposite was true in the Americans' data. The fact that the F0 range data was different for American and British women demonstrates that the interaction between pitch and gender can be extremely culture specific. Interestingly, the author also asserts that prosodic factors may be modified less consciously than segments are, as the speakers in the study were more aware of segmental differences than they were of prosodic ones.

At least some studies suggests that pitch and how it interacts with gender can vary across languages and across social groups within languages. However, skeptics may suggest that cross-cultural differences such as those Loveday reports could be attributed to physical differences between the subjects from the two cultures (especially if one is skeptical of the validity of measures of phonational range). To further explore the role of cultural expectation on pitch, Ohara (1992) studied the same speakers (same larynxes!) speaking two different languages, again languages with different gender role expectations, Japanese and English. The hypothesis was that native Japanese speakers would modify their pitch level when speaking Japanese versus when they speak English, i.e., females would raise their pitch when speaking Japanese, but not English. Six male and six female bilingual subjects in their 20s were recorded reading ten sentences in each language. The results support the hypothesis in that female subjects used significantly higher pitch when speaking Japanese versus English: "for females, the average frequency is at least 19 Hz greater when speaking Japanese than English" (p. 473). The pitch range was also greater for women in Japanese. No such differences were found for the men. Ohara concludes that women may be modifying their pitch to convey culture-specific expected femininity. The results provide evidence that some gender differences in pitch are not purely physiological. They are tempered by the fact that the tokens were read speech, and the sentences may not have been equivalent in English and Japanese—the English ones were from Ladefoged's *A Course in Phonetics* (Ladefoged, 1982), to demonstrate various intonations, and the Japanese ones were "translations," which meant the same meanings, not the same variety of intonations (Ohara, personal communication).

The same line of study was continued in Ohara (1999), where F0 was found to interact with gender and culture (and addressee, see below). The study consisted of five male and five female native speakers of Japanese who were fluent in English (NJ) and five female and five male native English speakers who were fluent in Japanese (NE) who were in their 20s, 30s and early 40s who were asked to leave a message on an answering machine in order to create a somewhat natural situation. Each subject was asked to leave four messages, one for a Japanese professor, a Japanese friend, an American professor, and an American friend in the appropriate language. They weren't given scripts but had to include certain items (the message was about borrowing a book and they had to leave their name and some other information). The means for the five pieces of information included in the message were averaged to get one average mean F0 for each message. Analysis of the results revealed that all of the female subjects (NJ and NE) used a higher F0 in Japanese than in English whether speaking to a friend or a professor (the variation was bigger for the native Japanese speakers). The author asserts that the results cannot be explained by physiology, language structure, or variance in emotional or physical (health) state and suggests that they are due to cultural differences in gender conceptualizations that affect pitch. In Japan, high pitch expresses a female role and leads to more positive perceptions of personality traits (Ohara has conducted some perception tests). High pitch is one important way female gender is performed in Japanese. Culture-specific pitch variations point to important non-biological factors in explaining women's use of pitch.

Males exhibit cross-cultural differences, too. Majewski et al. (1972) found that a group of 103 Polish males (ages 17-28) who were recorded reading a passage had an average speaking fundamental frequency (138 Hz) that was eight Hz higher than comparable data from American males (see Hollien and Jackson, 1973). The hypothesis that the difference was due to physical size differences between the two groups (the Americans were slightly taller and heavier) was rejected, as statistical analysis showed no significant relationships between SFF and any size parameter in the data. Thus, cultural differences (perhaps in gender roles) may be responsible for the differences. Furthermore, van Bezooijen (1995) suggests that a strong emphasis on masculinity (and low pitch) for Japanese males may explain some of the socially-based pitch differences between Japanese men and women (not to mention the lack of difference in her study between Japanese and Dutch women's pitch, see above). In this study, it was the concept of the ideal man that differed most between Japanese and Dutch cultures. She also speculates that this emphasis on masculinity in Japanese can explain Loveday's (1981) findings that Japanese men had lower top pitches than English men did. Thus, it may be the men that are doing more social manipulations of their pitch in some cultures to convey their gender. Regardless, the fact that both men's and women's fundamental frequencies vary cross-culturally in ways that apparently cannot be explained by physical size indicates that social norms and influences tell an important part of the story in describing voice pitch within the sexes.

1.4. Social Influences: Personal Identity/Strategies

1.4.1. Phonetic Effects of Gender Identity

Aside from the influences of the cultural or social groups speakers belong to, speakers' own personal identities or their version of femininity may affect their vocal linguistic strategies.

Biemans and van Bezooijen (1996) did not get the results they expected in their study exploring whether an individual's gender identity within a given culture has an effect on their pitch (rather than exploring differences in pitch and gender identity between cultures as wholes). They measured speakers' gender identity, i.e., "the extent to which gender stereotypical behavior, thoughts, or feelings are part of a person's identity" (p. 26), their average pitch, and their pitch setting, the difference between a person's average pitch and the lowest pitch he/she can produce. Subjects were 30 Dutch women, ages 20-45. Speech samples were semi-spontaneous speech in the form of dyadic conversations between subjects who knew each other about pre-arranged, provided topics. Gender identity was measured with a Dutch questionnaire which reflects beliefs about masculinity and femininity in the Netherlands and assesses traits and behaviors. Subjects also filled out the traits part of the questionnaire about their conversational partner. The hypothesis was that speakers with a more masculine gender identity would use lower average pitches and have smaller pitch settings. Analysis of the results showed that the speakers' gender identity scores on themselves did not differ much from the scores their conversational partners gave them, but there was only one significant positive correlation involving a pitch measure, that between masculinity and pitch setting which suggested that "masculine" speakers will use a pitch that is relatively far above their lowest pitch—

contrary to expectations. None of the 29 traits and behaviors correlated significantly with the two pitch measures. The hypothesis was not supported. The women were educated, which may be significant, or, it may be that perceptions of how women talk (here, either feminine or masculine ones) do not reflect reality (acoustic data) but rather are evidence of stereotypes that aren't accurate (cf. Henton's articles).

Biemans (1998) reports the results for both male and female Dutch speakers in dyadic conversations, and found that gender identity did not correlate strongly with various pitch measures (minimum and mean pitch and pitch range) for males, either. For both men and women, the only significant correlations, involving higher masculinity scores and higher minimum pitch, were the opposite of what the author expected.

1.4.2. Effects of the Interlocutor

Very little speech takes place without an interlocutor, and it is interlocutors who perceive our speech and make judgments about us. Furthermore, comparing speakers' pitch characteristics as they vary between different interlocutors is yet another way of isolating social influences, as speakers use the same physical apparatus whomever they speak to. There are two ways in which interlocutors could affect speakers' speech characteristics: Expectations about appropriate speech to a given class of people and level of accommodation, i.e., how much speakers' pitch characteristics become convergent.

1.4.2.1. Expectations Based on Characteristics of the Interlocutor

People may have ideas about what is an appropriate way to speak to someone with certain social characteristics. Aside from women being taught to talk like ladies (and men being taught to talk like men), people may be taught how to talk *to* a lady (or a man). For example, in her recordings of semi-structured play, in addition to finding the gender difference that fathers interrupted their children (ages 2-5) more than mothers did, Greif (1980) found that both mothers and fathers interrupted and spoke simultaneously with their daughters more than their sons. Brouwer et al. (1979) found very little evidence of gender differences among speakers in their study of interactions between ticket sellers and customers in a Dutch train station. However, they suggest that the factor explaining the usage of certain variables, all stereotypically associated with female speakers and considered markers of insecurity and politeness, is actually the sex of the addressee; both male and female speakers used a significantly larger number of words over all and significantly more diminutives, civilities, and hesitations when speaking to a male ticket seller. The authors recommend that, "Sociolinguists must start from the principle that the selection of the interviewer needs as much attention as the selection of the informants..." (p. 49). Chapter 8 in Giles and Powesland (1975) provides a review of previous work on the influence of personal characteristics of the receiver on a wide variety of different aspects of speakers' language and performance.

Several studies have found that characteristics of an interlocutor such as gender and status can affect a speaker's use of pitch. For instance, in her study of Dutch speakers conversing with friends, Biemans (1998) found that women and men used higher pitch with interlocutors of the opposite sex; men and women both used significantly higher median pitches in the opposite sex context, and the women had a significantly bigger pitch range (due to higher maximum pitches). Citing the frequency code as an explanation and perceptual studies on personality attributions, she speculates that "in mixed-sex conversations speakers aim at appearing less dominant and competent than in same-sex conversations" (p. 50). She posits as an alternative explanation that separate phenomena explain women's and men's behavior; women may be using a higher pitch to sound more feminine with men, and men may be attempting to lessen the interpersonal distance with their female interlocutors via an accommodation strategy (see next section). However, Edelsky (1979) found that more women used a rise-fall-rise intonation contour when answering female interlocutors when they were approached and asked questions in a public setting.

Ohara (1999), in addition to finding cross-linguistic differences in bilingual speakers of Japanese and English, found effects of the interlocutor's status, in particular for female speakers: All native female Japanese speakers used higher pitch with professors as opposed to peers in Japanese when leaving a message on an answering machine (it wasn't clear whether the professors were represented as male or female). The males did not show this pattern. In English, it could be the case that there is a pull toward higher pitch in order to be submissive or nonthreatening, and there is a pull toward lower pitch in order to be serious/adult/scholarly; these factors could lead to variation, but this explanation is purely speculative.

In her studies of Japanese, Yuasa (1999) didn't find effects of the gender of the interlocutor for her (male) speakers, but did find an effect of their familiarity with the speaker. Speakers used a narrower pitch range with interlocutors they did not know well.

Female and male American and Japanese speakers were analyzed in Yuasa (2000), where the same familiar/unfamiliar distinction was found for both male and female Japanese speakers. Furthermore, both American and Japanese women had larger maximum pitch excursions than their male counterparts (American females only spoke with a familiar female interlocutor, the researcher). Perhaps in Japanese, the politeness system suggested by Yuasa, where emotional expression via wider pitch ranges is constrained with unfamiliar interlocutors, masks or outweighs any effects that the gender of the interlocutor may have, or, they may not exist.

1.4.2.2. Accommodation to the Interlocutor

While Greif (1980) found gender differences only in the parents and not in the children in her study, Lieberman (1967) found that infants (a 10-month-old boy and a 13-month-old girl) adjusted their median F0 according to which parent they were interacting with. They used lower pitches when babbling with their fathers than when with their mothers (averages for boy: 430 Hz when alone, 340 Hz with father, and 390 Hz with mother; for girl: 290 Hz with father and 390 Hz with mother). As Lieberman's interpretation suggests, these (fairly young) infants, rather than adjusting according to what they thought was appropriate for each parent, were probably mimicking (or attempting approximations) of their parents' F0s.

This kind of mimicry points to another possible way that speakers may adjust their speech to that of their interlocutor, i.e., accommodate to them and speak more like they do. Swingle and Hope (1987) found evidence of convergence in percentage of time spent phonating and amount of pausing by speakers in an interview situation. Their data suggest that the interviewer and interviewee converge at the start of the interview and that by the end of the interview, the interviewer's vocalization was affected by the interviewee's. The interviewer was male and the interviewees were four males and five females, but either there was no effect of the gender of the interviewee or it was not discussed.

Sociolinguists have done significant work in exploring accommodation between speakers in conversation, although previous work focuses on vowel quality and other segmental and lexical aspects of speech. Bell (1984) develops a theory of accommodation, describing style changes within individuals as audience design, i.e., accommodating to addressees, typically to win approval. He asserts that intra-speaker variation mirrors and is derived from inter-speaker variation in that a speaker shifts styles in order to sounds more like another speaker who belongs to a different social/dialect group[5]. Coupland (1985) and Rickford and McNair-Knox (1994) provide good examples of the application of Bell's ideas to explain phonological variation between Cardiff English and African American Vernacular English with standard dialects. (See Chapter 4 for further discussion of motivations for accommodation.)

It is not always clear which category (expectation or accommodation) a given effect falls into, as Biemans (1998) points out. For instance, while there has not been a lot of research comparing loudness between men and women, one study by Markel et al. (1972) did study the effect of sex of subject and sex of experimenter on speaking intensity in dyadic conversations. Average intensity level was obtained by averaging peaks of intensity in the samples measured by a graphic level recorder. They found that speaking

[5] See Finegan and Biber (1994) for an alternative explanation, i.e., that register variation explains social group variation, rather than vice versa.

intensity indeed varies as a function of the sex of the subject and the sex of the experimenter. Men speak louder than women, and within each gender, people speak louder to people of the opposite sex. While the effect is clear, its explanation is not. It could be that each gender thinks that it is appropriate to speak more loudly to the other, or a partial explanation could be that women are accommodating to men's loudness (since men in general spoke louder than women in this study).

Previous research on interlocutor gender has found some evidence of the influence of both expectations of appropriateness and accommodation on speakers' speech characteristics. It remains to explore them more fully and to separate the two effects for pitch characteristics.

1.5. Conclusion

The research cited above demonstrates the vast array of potential candidates for causes of gender differences in pitch: anatomical and physiological differences, and three kinds of social influences, those based on group membership, those based on personal identity, and those due to effects of the interlocutor. In an attempt to isolate and elucidate the role of one particular kind of candidate, my experiments will focus on this last kind of social influence, the effect of the interlocutor. Previous research guided the design of my experiments, which will be described in the following chapter. The questions explored in the experiments are: If explored in a controlled experiment, what are the influences of interlocutor gender and status on female speakers' pitch? Are any effects found due to accommodation to the interlocutor? Two possible hypotheses regarding gender considered were that subjects would use higher pitches in some manner with male interlocutors, perhaps heightening femininity cues (cf. Biemans, 1998), or that they would use higher pitches with other females as a camaraderie politeness strategy (Lakoff, 1973, Tannen, 1984, Brown and Levinson, 1987). Regarding status, it was hypothesized that subjects would use higher pitches in some manner with professors as a deference strategy or to sound non-threatening (Brown and Levinson, 1987, Ohala, 1994). (See Chapter 4 for a discussion of different politeness strategies.) The results of my experiments will provide another piece in the puzzle as far as explaining female speakers' pitch patterns, and finding the dividing line between physiological and social explanations, at least within one specific culture.

The underlying hypothesis is that women use pitch as one of many ways of constructing their gendered identities, as one of many symbolic tools signaling their views of themselves, others, and their relationships to others and the world. Many scholars have recognized and studied the central role that language in its many facets plays in constructing gender. Ochs (1992) asserts "Gender ideologies are socialized, sustained, and transformed through talk, particularly through verbal practices that recur innumerable times in the lives of members of social groups" (p. 336). In this discussion of how the linguistic strategies of mothers serve to construct female roles and shape how women are perceived, Ochs articulates a model of the relationship between language and gender, suggesting that the relationship is mediated by language's connections with other social constructs and pragmatic meanings (such as stances, acts, and activities). Linguistic forms are used to perform pragmatic work that is differentially distributed and expected according to gendered social identities. She cites tag questions in English as being tied both to female speakers and to the marking of stances like hesitancy and particles in Japanese that are associated with delicateness as well as with female speakers as examples of the multi-tiered connection between language and gender. Furthermore, she explicitly recognizes the use of pitch as a non-referential index of gender (p. 339). To do justice to the complex relationship between linguistic devices and gender, Ochs advocates a functional or strategy-based account of gendered speech that focuses on activities and then speech. While my design was constrained by the need for a controlled experiment and good recordings, I did try to elicit certain kinds of speech acts and create activities within the interview situation (see Chapter 2).

By focusing on female speakers in order to explore how gender can be socially constructed via phonetic cues, I am not equating gender and language studies with women's language studies, which is tantamount to seeing male speech as normative and women as a special case (see Eckert and McConnell-Ginet, 1992, for a discussion of this idea). The data from my experiments can contribute to what needs to be a multi-faceted and cross-disciplinary exploration of gender construction which of course studies both male and female speakers. Furthermore, Eckert (2000) argues that "the primary importance of gender lies not in differences between male and female across the board, but in differences within gender groups. In developing patterns of behavior, in assessing their own place in the world, and in evaluating their progress, people orient above all to their own gender group" (pp. 122-3). As speakers create their gendered identities via speech and other symbolic behavior, they do so in comparison to others in their own gender group, selecting amongst and creating options to signal what kind of female or male they are. My study can help elucidate one aspect of how women of a certain culture construct their gendered identities and do so in a way that varies according the social situation they are in (i.e., depending on the interlocutor and the topic of conversation).

In her book on different kinds of discourse analysis, Schiffrin (1994) provides a more general framework for my study, alluding to John Gumperz' ideas: "Speakers are members of social and cultural groups: the way we use language not only reflects our group-based identity but also provides continual indices as to who we are, what we want to communicate, and how we know how to do so" (p. 102). Speakers' language use reflects their membership in a gender category, but within that identity, their linguistic and paralinguistic variations are tools used to communicate their attitudes toward their interlocutors and their views of themselves in relation to others.

Chapter 2. Methodology

2.1. Review of Past Methods

To control for content and to facilitate laboratory recording, many studies measuring fundamental frequency in women and men use read speech, asking subjects to read one or more typically neutral passages. To varying extents, read speech can lead to data that is rather far-removed from everyday speech, i.e., speech that reflects our gendered identities and politeness strategies and is the basis of impression formation. Read speech is not the same as spontaneous speech, and people may have different styles of reading aloud as well as different comfort levels or even competency levels; analysis of read speech may thus be most telling about a person's reading style or reading ability. These aspects of read speech make it less than ideal for use in studying aspects of speakers' pitch if the goal is to study how they use pitch in everyday life. Some studies that have used read speech have compounded these problems in various ways. For instance, Stoicheff's (1981) subjects were instructed to read a passage as if they had an audience of 25 people, which might lead to use of a particularly formal register and more than conversational loudness.

Loveday (1981) was interested in cross-cultural pitch differences. Because he was interested in pitch differences in politeness formulae, he had subjects read a part in a dialogue that included several such expressions. A possible problem could be that, just as people may vary in their reading styles and abilities, they may vary in their acting abilities. Dialogue readings could be well-acted but stylized, or they could be extremely unnatural if a person isn't able to empathize with the situation in the scripted dialogue.

Despite such problems, there are some advantages to using read speech. Positive aspects of eliciting read speech consistently across subjects, such as in de Pinto and Hollien (1982), Linke (1973), Künzel (1989), Günzberger (1996), van Bezooijen (1995), and Henton (1989, 1995), are that content (and even register) is controlled for and that high-quality recordings are possible (e.g., using good equipment in a sound booth). One way of dealing with the variety of methods for eliciting material for pitch analysis is to compare within methodologies, i.e., read speech can be compared with read speech (especially if the content of the passages is similar), etc.

Other studies measuring aspects of pitch did not use (only) read speech. Some researchers designed clever methods to try to control for content but still obtain somewhat natural speech. Awan and Mueller (1992) asked their subjects to describe a picture, which would control for topic and task/register (and even possibly elicit some of the same lexical items from different subjects), and Ohara (1999) asked subjects to leave a message on an answering machine with certain pieces of information included. The answering machine format was also useful in that the subjects could be told who they were leaving messages for, i.e., the addressee could be varied. Extemporaneous speech can also be elicited in the form of spontaneous monologues, as in Shevchenko (1999). However, amongst the studies using extemporaneous speech, some methodological problems are apparent. It is unclear whether Shevchenko let subjects speak on a topic of their choice or whether a topic was provided. If they all spoke on different topics, their choice of topic may well have had an effect on aspects of their speech. Also, in Hollien and Jackson's (1973) study of young men (in addition to using a read text), they asked their subjects to speak on a provided topic, but used four different topics. Again, different topics may well have inspired different uses of pitch. They also told their subjects that their voices were being studied so that they would not be self-conscious about constructing their speech (content), but this information may have made them self-conscious about their voices. However, all of these kinds of spontaneous speech are closer to real-life speech than read speech is and can still be recorded in a laboratory setting.

Perhaps the most natural speech samples are those from conversations, but using conversations involves a host of other difficulties. Obtaining speech samples in a natural setting is extremely difficult, as people tend to act unnaturally when they know they are being recorded, and outside of a laboratory, it is very difficult to make high-quality recordings. Furthermore, there is less control over topic and other situational factors. Some researchers have used data from conversations, trying to strike a balance between maintaining naturalness and controlling as many variables as possible. For instance, Biemans and van Bezooijen (1996) asked women who were acquaintances to discuss a prescribed topic in one of the speakers' home, thus controlling for gender, familiarity, and topic, and obtaining fairly natural samples. (However, there was more than one prescribed topic and the different topics may have induced different uses of pitch.) Yuasa (1998, 1999, 2000) also used data from conversations in her studies of pitch range in Japanese and controlled familiarity and gender of the interlocutor, but did not always control for topics. Furthermore, when interlocutor gender was manipulated, half of the subjects spoke with female interlocutors and half with male interlocutors; individual subjects did not speak with interlocutors of both genders.

Another common problem in these pitch studies is the use of a small number of subjects. Such studies run the risk of capturing idiosyncrasies of individuals. There is probably a lot of individual variation in pitch use; averaged data from a large number of subjects could be more useful.

2.2. My Experimental Design

The goals of my experimental design were to control for as many factors as possible and obtain high quality data and at the same time have the data be as natural as possible. Both goals suffered in the compromise, but neither was ignored.

2.2.1. Subjects

Subjects were 16 UC Berkeley undergraduates, 12 females and 4 males, solicited via email sent to introductory linguistics classes and flyers posted around campus[6]. In order to ensure that the subject pool was as homogeneous as possible, the email and flyers specified that they had to be Caucasian, heterosexual, 18-22 years old, native speakers of California English with no hearing loss. (One subject was 23.) Since I did not plan on exploring the effects of race, age, dialect, or sexuality, I needed to control for these factors. As a way of explanation as to why they were being recorded talking to four different people, subjects were told the study was on interviewing techniques. It was hoped subjects would feel less self-conscious and be less apt to monitor their speech and act unnaturally if they thought it was their interlocutors' speech that was to be analyzed.

2.2.2. Experimental Conspirators

As I wanted to explore the effects of both gender and status of the interlocutor, there were four experimental conspirators, a male playing the part of a peer of the undergraduate students, a female playing the part of a peer, a male professor, and a female professor. The peer conspirators were two UC Berkeley graduate students, a male and a female, in their mid-twenties who introduced themselves as undergraduate students in the recording sessions, and the professor conspirators were two UC Berkeley Linguistics professors, a male and a female (ages 59 and 46), who introduced themselves as such. All conspirators were Caucasian.

Personality of the interlocutor was not manipulated, but was controlled for in that all subjects spoke with the same four interlocutors. They were all skillful conversationalists and friendly individuals. Subjects could have been reacting in part to the individual personalities of the interlocutors, but this possibility could have been removed only if there were a large number of each kind (male vs. female and peer vs. professor) of interlocutor. Another possible problem in the design is that age and status could have been confounded; the professors were older than the peers. Once again, the only way to avoid this problem would have been to have very large numbers of interlocutors of different age and status combinations, which would have been logistically near impossible. However, both professors were significantly older than the peers, so if age was a confounder, it at least confounded similarly for both professors. Furthermore, it may be that the prototypical professor is not extremely young, so this situation could have helped elicit prototypical "professorial" scenarios.

---

[6] Only female speakers' data will be analyzed in this project.

Interlocutors were extensively coached as to the format of the interviews and provided with a list of questions and possible follow-up questions to ask each subject. The interviews were each preceded by an introduction and a brief period of casual chatting about subjects' majors, interests, etc.

2.2.3. Recording Procedures

The recordings took place in a sound-treated room in the Phonology Lab at UC Berkeley in the spring of 2001 over the course of a few months. After signing informed-participation consent forms, subjects were taken to the sound booth. The subject and conspirator sat in fairly close proximity at a table with their chairs partially turned towards each other. Subjects and conspirators wore head-set microphones, which I adjusted to make sure they were positioned correctly. Having the subject and interlocutor sitting face-to-face did compromise the channel separation in the recordings, i.e., at times the speech from one speaker is audible in the other speaker's channel, but it was imperative that the pairs be face-to-face for the sake of conversational naturalness. The channel separation issue was dealt with in the acoustic analysis (see section 2.3 below).

I was present for all four interviews for four of the subjects (S1, S2, S13, and S14, but only two of these were females analyzed here) before deciding that my presence might be slightly impeding the naturalness of the conversations. All recordings took place in the morning hours. The conversations were digitally recorded at 48 kHz, 16 bits per sample. Each subject participated in two recording sessions, each involving two approximately 15-minute interviews. The subjects were interviewed by the conspirators in a constant order, by the female peer and then the male peer in session one and by the female professor and then the male professor in session two. After the first conversation within each session, the subject remained in the sound booth while the first interlocutor they spoke to left and got the second interlocutor. After the second recording session, they filled out two questionnaires (PAQ and Feminism Scales), were debriefed, and signed debriefing and data release consent forms. Subjects were paid.

2.2.4. Interview Topics and their Sequence

Much of the previous work exploring pitch in conversations either ignored the problem of topic of conversation or dealt with it by controlling for it by keeping it relatively constant. Given the difficulties of fully controlling topics in relatively natural conversations, in my study, I attempted to explore the influences of conversational topic by manipulating it as I explored interlocutor influences. It was my hope to manipulate topic in a "gender-savvy" manner, so that the topic variations would be meaningful in an experiment exploring how gender is constructed in different contexts.

Thorne's (1993) description of elementary school kids' interactions demonstrates how gender is highlighted and muted according to context, i.e., the activity and the location. At certain times in school, such as working together on a project in class in small groups, gender is not salient, but at other times, such as in cross-gender chasing on the playground, it is brought to the forefront and constructed as both antagonistic and dualistic. Thus, gender "is not only a category of individual identity and the focus of symbolic constructions, but also a dimension of *social relations and social organization*" (pp. 158-9, author's emphasis).

Because gender is constructed and given meaning throughout all interactions, Thorne challenges researchers to examine gender in all its social contexts rather than focusing on sweeping generalizations and binary oppositions (pp. 108-9). My research aims to examine how gender expresses individual identities as they are constructed (variably) across different social contexts. The social contexts created in my experimental conditions are somewhat artificial due to the need to isolate certain variables and control for others, but many aspects of natural face-to-face interaction are preserved, and it can serve as a starting point for further research.

Subjects' speech in response to three topic genres was studied. The order of topics was held constant, with a brief period of chatting followed by topic A, then B, and then C. To avoid the unnaturalness and possible effects on the discourse of telling the same stories repeatedly, semi-equivalent topics were constructed for each genre. However, with the aim of eliciting natural conversations, separate topics were constructed for the peer (topics three and four) and professor (topics one and two) interviewers within each topic genre. The hope was that phonetic information measured would thus be less likely to be due to embarrassment and obscure the effects of the social parameters under study. Furthermore, to avoid confounding the specific topics with the social variables of the interlocutors, the topics were randomized in a controlled manner. They were rotated so that the four topics within a genre were used equally and used an equal number of times by a given peer or professor interviewer. For example, Subject 1 had topics A1, B1, and C1 with the female peer and topics A2, B2, and C2 with the male peer, while Subject 2 had topics A2, B2, and C2 with the female peer and topics A1, B1, and C1 with the male peer. All possible combinations were used equally (see Appendix 2).

Topic Genre A

Topic genre A was designed to be "gendered," i.e., to elicited subjects' gender identities. To this end, the questions were designed to cause the subjects to think about their gender roles and identities so that they would be within that "frame" when answering. The following were the topic genre A questions:

A1: [The experimental conspirator tells a loosely-scripted story about how his/her car broke down that morning and would need expensive repairs. The story was designed to elicit sympathetic responses from the hearer (subject), which were hypothesized to be gendered, particularly in their pitch and voice quality characteristics.]

A2: Do you like being a female/male? Specifically, are there times when you feel it is an advantage or disadvantage or that it brings about or interferes with any opportunities?

A3: Do you have a boyfriend/girlfriend? Tell me all about him/her. If no, what would your ideal man/woman be like?

A4: Did you rush (a sorority/fraternity)? If so, tell me about your experiences and the friends you made. If no, did you live in the dorms your first year? Is that where you made a lot of your friends? What do you do when you hang out with just your girl/guy friends?

Topic Genre B

This series of questions were designed as a control group, i.e., to be mundane and at the very least, (relatively) non-gendered.

B1: What classes are you taking this semester, and with what professors? Where and when do they meet? What are the course requirements?

B2: How do you get to school? (What streets do you or your bus take?)

B3: What all did you eat yesterday?

B4: Describe the layout of your house—where are the rooms in relation to each other and how are the rooms set up?

Topic Genre C

Topic genre C aimed to create a situation where the subject would be forced to commit some kind of a face-threatening act (FTA) (see Brown and Levinson, 1987, for definitions of face and discussions of how it is threatened). The FTA was drawn by having the peer interlocutors elicit an opinion from the subject and then disagree with it and press the subject further. For the professor interlocutors, topic C1 involves the professor purposely making a mistake in a brainteaser problem, and topic C2 involves a role-play where the subject is to point out a grading error on the professor's part. The hypothesis considered here is that potential phonetic markers of femininity might be heightened when the subject is forced to verbally do something aggressive, here, disagree with or correct the interlocutor.

C1: I'm going to ask you do a brainteaser. I'll do one first to give you an example: Jack Axe charges $5 to cut a wooden log into two pieces[7]. How much will Jack charge to cut a log into four pieces? [(*Answer: Jack will charge $15 since it requires three cuts to make four pieces.)* Then, you (the conspirator) will ask them the following brainteaser: Art Conn bought a used car for $600 and sold it to Hardy Pyle for $800. He later bought it back for $1000 and resold it for $1200. Did Art make any profit and if so how much? *(Answer: At first glance it appears that Art Conn made a profit of $200, however this is*

[7] These two brainteaser questions were taken from the game MindTrap, and are the property of MindTrap Games, Inc. copyright 1991.

not the case. Art made a total profit of $400 since he made $200 profit each time he sold the car.)]

C2: Role-play. Let's pretend that I'm your professor and I have made an error grading your exam—Say I marked a question wrong that you believe you answered correctly (you circled "a" and that is the correct answer), and if that question were counted as correct, you would get an A minus rather than a B plus on the exam. You want to bring this to my attention so you come to my office. What do you say to me?

C3: Are you religious? If no, why not? Do you believe in God? Why not, or if so, why don't you practice a religion? For yes answers (and for no answers, too, if there's time): What do you think about homosexuality? In particular, do you think that homosexual people should be allowed to legally marry and/or adopt children? (You can also ask about prayer in school.)

C4: What is your opinion on affirmative action in general? Do you think it should happen in admissions at UC Berkeley?

2.2.5. The Questionnaires

When Biemans and van Bezooijen (1996) tried to find a relationship between pitch measures and the gender identities of their Dutch subjects, they found very little. The one significant positive correlation in their data was between masculinity and pitch setting, the distance between average pitch and the lowest pitch the subject can produce, and this finding was counter to their hypothesis. Without having strong expectations of positive findings, for the sake of completeness, I distributed a gender identity survey to my American subjects.

The gender identity survey I chose to use was the Personal Attributes Questionnaire (PAQ), described in Spence and Helmreich (1978). This survey measures the personality traits associated with masculinity and femininity, i.e., "stable internal characteristics of the individual" rather than the sex role behaviors that other surveys measure (O'Grady, Freda, and Mikulka, 1979, p. 216). These traits are described as "independent clusters of socially desirable attributes commonly believed to differentiate the sexes" (Spence and Helmreich, p. 115). The of use positive traits and the concepts of masculinity and femininity underlying these surveys led to the development of three scales, one for masculinity, one for femininity, and a quasi-independent polarized masculinity-femininity scale. Masculinity and femininity are on separate scales as they are considered to be orthogonal to each other, rather than in a bipolar relationship; masculine traits are associated with a sense of agency and feminine traits are associated with a sense of communion, i.e., instrumental traits and expressive traits, respectively. In the interests of time, the short (24-question) form of the PAQ was administered.

In addition to exploring the possibility that gendered personality traits could lie behind any pitch pattern differences found between subjects, I wanted to test whether a subject's belief in feminism might correlate with her pattern of adjusting or not adjusting her pitch patterns according to her interlocutor. To this end, I adapted the up-to-date and comprehensive Feminist Perspectives Scale developed by Henley and colleagues, described in Henley, Meng, O'Brien, McCarthy, and Sockloskie (1998) and Henley, Spalding, and Kosta (2000). These surveys include questions for five attitudinal feminist subscales, liberal, radical, socialist, cultural, and womanist, and a behavioral scale. Again in the interests of time, I used the short form and did not include the behavior scale or the questions for the womanist scale, as these questions refer to the points of view of women of color, and by design, all of my subjects were Caucasian. I also excluded the socialist feminist questions, as they were less relevant. The survey includes questions to create a conservatism index, which I did include. As previously mentioned, subjects were given these questionnaires after the second recording session but before they were debriefed. They were instructed to answer quickly and not to go back and change any answers.

2.3. Acoustic Analysis

2.3.1. Experiment One: Expectations due to the Addressee

In this experiment, several aspects of subjects' pitch were measured to test whether or not they were affected by the factors manipulated in the experimental design. To ensure that there would be little interference from interlocutors and that the measurements would be reflective of the speech data, solid minutes of data were prepared for each subject with each interlocutor for topic genres A and B, and for the first minute of the interview. (Topic C will not be analyzed here.) Laughter, pauses, extraneous noises, interlocutor speech, and overlapping speech were cut out so that only solid speech from the subject remained. I did not excise voiceless segments, but the pitch algorithm did not make measurements for them. The first minute of the interview contained introductions and chat, but contained varying amounts of topic A, so these portions were analyzed separately from the topic A and B minutes. For some subjects, solid minutes of data were unavailable, i.e., when interlocutor turns, pauses, etc. were omitted, that portion of the interview was less than one minute of solid subject speech. For these subjects, the data available was used, as long as it exceeded 40 seconds. (See Nolan, 1983, cited in Henton, 1995, for the suggestion that at least 40 seconds of consecutive speech are needed to make reliable pitch measurements.)

It was assumed that short-term pitch effects such as intrinsic pitch of vowels and consonantally-caused pitch perturbations in vowels would not be confounders, as the samples were long enough that such small effects would not skew the results. These phenomena affected all of the data and thus would not affect any one sample unduly. Furthermore, any declination present would be relatively uniform, as the kind of speech (extemporaneous conversation) was uniform throughout the interviews.

Measurements made and studied were: median pitch, standard deviation of pitch, and 80% pitch range (90^{th} percentile minus 10^{th} percentile). I chose these measures as previous research suggested they might be fruitful, i.e., relevant to gendered aspects of pitch (see literature review in chapter 1). As a general measure of pitch level, median pitch instead of mean pitch was considered as pitch data has been found to have a slightly non-normal distribution, being positively skewed. As more pitches occur in the lower ranges, the median is more reflective of the data than the mean, which may be unduly influenced by the fewer but more extreme higher pitches in the distribution (Jassem, 1971, Graddol, 1986). Standard deviation of pitch was measured as an indicator of pitch variability (Henton, 1995, Tielen, 1992). Eighty-percent range was measured instead of the full pitch range, as the very extreme pitches tend to be most sensitive to measurement errors (Graddol, 1986).

Measurements were made using the autocorrelation method in Praat software (see Boersma, 1993, for a description of the algorithm). A unique combination of settings within the Praat algorithm was created for each subject based on what gave the most accurate results; I experimented with settings until I eliminated/reduced spurious or suspicious measurements, and then these settings were used consistently, i.e., the same settings were used for all measurements made on data for a given subject. See Appendix 1 for a listing of these settings for each subject. Histograms of the pitch data matched these findings reported in the literature on pitch measurements. They tended to be positively skewed, with a very small number of many different high frequencies occurring. Looking at the histograms as well as their summary statistics and finding that they are the shape reported in the literature and do not have apparently spurious measurements also reaffirms my faith in the pitch measuring algorithm.

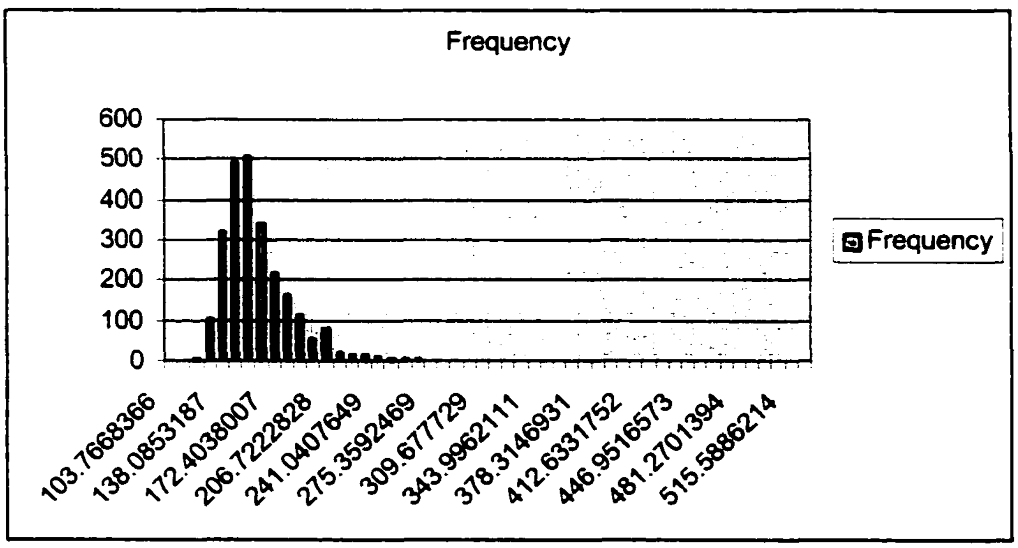


Figure 1. Subject 12, histogram of solid first minute of pitch data with male professor. 2491 data points.

Histograms of the data also tell me that the measurement of the full pitch range (without

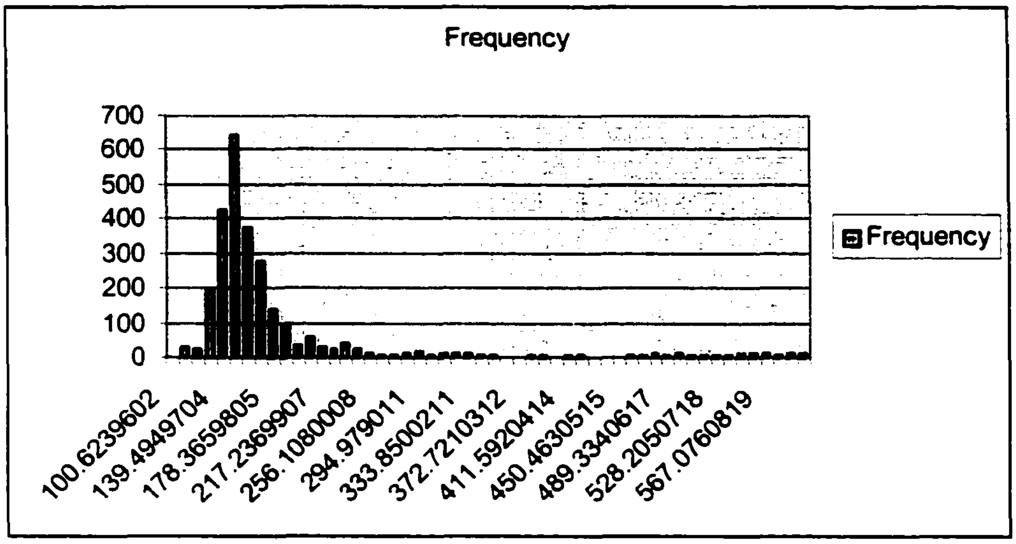


Figure 2. Subject 5. histogram of solid first minute of pitch data with male peer. 2663 data points.

throwing out the extreme pitch measures below the tenth and above the ninetieth percentile) are indeed unreliable, just as the literature suggests (Graddol, 1986). The measurements, especially those above the ninetieth percentile, are very high and quite probably spurious. They were thus not used in the analysis and only the 80% pitch range was considered.

Median pitch was measured and reported in Hertz. However, Henton (1989, 1995) and others comparing male and female speaking ranges concluded that due to the logarithmic nature of pitch perception, it is misleading to use a linear scale such as Hertz to compare intervals at different frequencies. A more accurate comparison can be made using a logarithmic scale such as the semitone scale. Psychoacoustic research, however, has shown the picture to be still more complicated in that frequencies below 500 Hz seem to be not entirely logarithmic. The ERB (Equivalent Rectangular Bandwidth) scale, a psychoacoustic scale, takes these findings into account. (See Moore, 1997, for a description of the auditory processes it takes into account.) Thus, range and standard deviation measures are reported in ERBs. While it may have been acceptable to use Hertz for my intra-sex (and intra-subject) comparisons, since the ERB scale reflects human perception, it is appropriate and safest given that the ranges and standard deviations being discussed all occur at slightly different levels of pitch. Furthermore, my data will be in a form where it can easily be used by other researchers measuring different women and those studying male speech and making cross-gender comparisons.

2.3.2. Experiment Two: Accommodation to the Addressee

This experiment explored whether or not participants in a conversation are affected by each other in their use of vocal pitch. The set of interviews[8] was the same as in experiment one, as were the pitch algorithm used and the measurements taken (median pitch, standard deviation of pitch, and 80% pitch range). However, the questions being explored were different, so the data were prepared differently. In this set of experiments, I endeavored to find out whether or not speakers accommodate their pitch to each other in any manner. I wanted to explore whether or not pitches converge in interviews and if so, in what manner. Because accommodation of pitch is relatively unexplored, a clear definition of it and how to measure it were not already available. Two kinds of accommodation were hypothesized to exist. It may be the case that a speaker adjusts her or his speaking voice so that it is more like an interlocutor's voice in a general or global manner, e.g., raising or lowering the tessitura/overall level of pitch or increasing or decreasing the variability to match the other person's voice to some extent. However, it may additionally or instead be the case that a speaker mirrors or mimics an interlocutor's pitch level or variability at certain points in the conversation, i.e., that the accommodation happens in "real time" and the speakers' pitch goes up and down together. Measuring moment-by-moment "across-time" accommodation is problematic because of the possibility of a phase difference. If speaker B is accommodating to speaker A in a conversation in real time, there could be a slight delay in the imitation that speaker B attempts, as he or she has to perceive and process the pitch patterns and then adopt them to whatever extent desired. This processing could happen somewhat instantaneously, or it

[8] The interviews between the peer interlocutors and S1 and S2 and between the female peer interlocutor and S4 were not used due to poor recording quality for the interlocutor channel in these interviews.

could cause a delay so that the imitating pitch pattern occurs 10, 20, or more seconds after the original. This possibility was taken into account in the preparation of the data described below, but it is difficult to adjust for it as the magnitude or even the presence of a phase difference is unknown, but entirely possible.

Another possibility that needs consideration is that either one or both people may be accommodating within a conversation. It could happen that speaker A is unaffected by speaker B's pitches, and her or his pitches reflect other sociolinguistic attributes of the conversation such as those described in experiment one (e.g., gender and status of the interlocutor, topic of conversation, etc.) as well as physiology. But, within this same conversation, speaker B is affected by speaker A's pitches; speaker B accommodates to speaker A but not vice versa. Another possibility is that both speaker A and speaker B are affected by the other's pitches. However, in practice, it may be extremely difficult to distinguish between "one-way" and "two-way" (i.e., mutual) accommodation, as both will involve some kind of a convergence of pitches. Furthermore, for mutual across-time accommodation, an additional obfuscation arises in that both speakers could be reacting to the same "trigger" if their pitches have the same pattern in an a given instant. They could both be reacting to the topic or any other aspect of the conversation that they both have access to or could both be getting excited or animated about the same thing. While this situation is indeed possible, it is unlikely that it could explain accommodation that persists across a conversation for any length of time. Despite all of these acknowledged difficulties, the question of whether pitch accommodation occurs and if so how, needs to begin to be answered in some fashion. My methodology and analysis can serve as a starting point to a line of inquiry.

In order to measure any accommodation occurring over the course of an interview and to get global, overall measures for entire interviews, I needed measurements from throughout the 15-minute interviews and for both of the speakers, and still had to excise interlocutor speech, overlaps, and silences. Speech samples where pitch measurements were made were uninterrupted sections of speech from one speaker that were two seconds or longer. To this end, a series of scripts were run on each channel of the interviews.

The first step in processing the data (the .wav file for each channel of each interview) was to run a script on it in Praat called "find silence[9]." A few files were unusable (S1 and S2 with the peers, and S4 with the female peer) due to poor recording quality on the interlocutor channel; these interviews were not included in the analysis. The "find silence" script used a silence threshold of 55 dB and considered silences longer than 0.2 s. to be pauses. Its output was a text grid within Praat with alternating sections of speech and silence labeled as such. The second step was to run a Perl script, "find overlap," on a pair of the text grids created from the first script (i.e., the text grids for the subject and interlocutor channel from the same interview). This script found the places in the two channels of a conversation where both channels contain speech and labeled them as overlap in both channels' text grids. Its output was two text grids with sections of speech, silence, and overlap all labeled. The third step was to run the "extract pitch" script within Praat on each paired .wav file and text grid with speech, silence, and overlap labels. This script extracted pitch data from the portions of the .wav file labeled as speech. The same settings within the pitch extraction algorithm developed for each speaker and used in

[9] I am indebted to Ronald Sprouse for writing these scripts.

experiment one were used, and similar appropriate settings were found and used consistently for each of the four interlocutors, as well (see Appendix 1). The autocorrelation pitch extraction method requires a minimum pitch duration of the length of at least three pitch periods; the minimum durations used were thus 0.1, 0.06, and 0.03 seconds for the minimum pitches (within 1 Hz) of 30, 50, and 100/125 Hz, respectively. The algorithm did not make measurements on portions of the .wav file labeled as speech that were less than two seconds long, but all of these intervals added together typically amounted to less than one second of speech per interview. The output of this third script was a text file of the pitch data, i.e., the median pitch, standard deviation of pitch, and 80% pitch range (the 90th percentile minus the 10th percentile) for each of the non-overlapped portions of speech as well as its start time, end time, and duration (start time minus end time), for both speakers throughout the interviews. ERBs were used for all measures, including medians, as male speech (that of two of the interlocutors) was measured in addition to female speech.

In order to measure the pitch of the speakers and interlocutors across time, the speech portions had to be grouped according to where they fell within the interview. In order to create a rough estimate of averages of measures across time, the speech sections that fell within one-minute-long intervals were thus grouped together. However, to lessen effects on the averages of where the dividing line between minute-long intervals fell, overlapping chunks were created. For instance, the first chunk of speech portions was from zero to 60 seconds, the second one was from 30 to 90 seconds, the third from 60.01 to 120 seconds, the fourth from 90.01 to 150 seconds, etc. The use of overlapping time chunks is non-ideal for statistical analysis, but gives a more accurate measure of the pitch patterns in a conversation. The smoothed averages help to minimize the effects of the boundaries of the minute-long chunks, but makes computing confidence intervals for the correlations described in Chapter 3 impossible.

These non-overlapped sections of speech in which pitch measurements were made and which were then grouped according to which minute they fell within were of various lengths, as the length of the speech section was determined by how long speech occurred with no pauses greater than 0.2 second or overlapping speech, per the application of the "find silence" and "find overlap" scripts. To obtain accurate averages within the minute-long intervals, i.e., to avoid measurements from shorter speech portions "counting" as much as longer speech portions, weighted averages were created. Within SPSS, for each of the three measures, the measurement for a given speech portion was multiplied by the portion's duration and then divided by the total duration of all of the speech portions of the minute-long interval in which it fell. These numbers were then summed for each minute-long interval. This weighted average will be referred to as "score." For the measurement of across-time accommodation, scores were obtained for each of the overlapping chunks. For the measurement of general, or global accommodation, a score for an entire interview was obtained by multiplying each measure within a speech portion by that portion's duration and then dividing this number by the total duration of all the speech portions for the entire interview. The process was parallel to that for across-time accommodation, but instead of dividing by the total duration of all the speech portions for each minute-long chunk, the division was by the total duration of portions for the whole interview. After the division by duration, these figures were summed within each interview/time chunk to obtain a score for each time chunk or each interview for each of the three pitch measures:

$$\text{Weighted Average for Global Accommodation} = \frac{\sum(\text{Portion's Measure} \times \text{Portion Duration})}{\text{Total Duration of Portions within Interview}}$$

$$\text{Weighted Average for Across - Time Accommodation} = \frac{\sum(\text{Portion's Measure} \times \text{Portion Duration})}{\text{Total Duration of Portions within Time Chunk}}$$

Thus, for global accommodation, one score for each measure for each speaker was obtained for each interview, and for across-time accommodation, one score for each measure for each speaker was obtained for each minute-long time chunk. Depending on the length of the interview, the number of overlapping time chunks ranged between 14 and 39. The total duration of speech portions within time chunks also varied greatly. A few minute-long chunks only had a few seconds of data, but most had from between 15 to even over 40 seconds of solid speech data. Borderline cases of correlation/accommodation were examined to make sure that a chunk with only a few seconds was not responsible for an erroneous result.

These weighted averages (scores) were analyzed in several ways. No pre-existing definition of pitch accommodation or methodology for measuring it was available, so various comparisons were attempted for both global and across-time data. Firstly, a normalized score was created for each measure for each time chunk (and for each entire interview), as raw weighted averages could not usefully be compared between speakers and interlocutors. The raw scores are reflective of individual speakers' voices, and there are very large inter-speaker differences both within males and females and especially between them. Particularly when groups of speakers are grouped together (e.g., when grouping the four interviews with a given subject together), the use of raw scores would obscure any patterns that might occur. The normalized scores were based on within-speaker comparisons, using each speaker's own median for each measure as a "yardstick" for comparison. A median (50^{th} percentile) was calculated for each measure for each of the 12 subjects and four interlocutors using all available data for that speaker[10]. For each subject, all data from all four interviews were used, and for the interlocutors, all data from all 10-12 interviews were used. For each measure (median pitch, standard deviation of pitch, and pitch range) and for each time unit (either minute-long time chunk or entire interview), the normalized score was calculated by subtracting this within-speaker median from the speaker's score and then dividing by the within-speaker median:

$$\text{Normalized Score} = \frac{(\text{Raw Score} - \text{Median})}{\text{Median}}$$

The normalized scores described above can indicate whether and how speaker-hearers respond to the pitch characteristics of their interlocutor, be it their median pitch, SD of pitch, or pitch range. However, it may be the case that what speaker-hearers respond to is the movement or change of a given pitch measure. For instance, instead of responding to an interlocutor's use of 200 Hertz, a speaker may respond to an interlocutor's increase or decrease of ten Hertz. In addition to comparing normalized scores between speakers and interlocutors, it could be important to measure the *rate of change* of the (normalized) scores across time, as this change may be what dovetails when accommodation occurs within a conversation. To test this idea, a score minus score was calculated and analyzed for the across-time data, i.e., a delta of the scores across time was calculated. For example, for each speaker and for each measure for all the interviews, the difference

[10] The "raw" (unweighted) data from the speech portions was used to calculate the within-speaker medians.

between the normalized score for time chunk two (30-90 seconds) and that for time chunk one (0-60 seconds) was computed, then the difference between the score for chunk three and chunk two, etc.

Just as in experiment one, pitch measurements were made from data with pauses and overlap excised. However, unlike in the first experiment where the data were prepared by hand, since the amount of data being considered in the accommodation experiment was much larger, the scripts described above were used to remove the pauses and overlaps. As a result, the data for the accommodation experiment were less "pure" in that laughter and extraneous noises were not found and eliminated. The pitches can thus be understood to be reflective of both linguistic and paralinguistic aspects of the conversations. However, it was often the case that extraneous noises and laughter occurred while the other person was speaking and, as overlapped data, was eliminated.

The above procedures were followed consistently and have been reported comprehensively. Despite the shortcomings of the methodology, it can serve as a starting point for the investigation of accommodation.

The data prepared as described above for experiments one and two were examined for patterns and statistically analyzed; this analysis is presented in Chapter 3.

Chapter 3. Results

3.1. Experiment One: Pitch Differences Based on Expectations Due to the Addressee

3.1.1. Pooled Data

Data from all 12 female subjects were analyzed together to find any general patterns. Using SPSS software, the results of pitch measurements on the 40-to-60 second solid-speech excerpts for each subject with each interlocutor for the beginning of the interview and for topic A (genderized) and B (mundane) were compared. The measurements of median pitch, standard deviation of pitch, and 80% pitch range (90th percentile minus the 10th percentile) were used as dependent variables to test the effects of interlocutor gender and status in a series of one-way Analyses of Variance (ANOVA). ANOVA determined that 80% pitch range was significantly larger with female interlocutors compared with male interlocutors ($F(1, 73) = 6.179$, $p < 0.05$). Means were 1.71787 ERBs (SD = 0.54673) with female interlocutors and 1.44597 ERBs (SD = 0.36970) with males, and medians were 1.505 ERBs and 1.38 ERBs with females and males, respectively. In the following boxplots, the horizontal solid line represents the median, interquartile ranges containing 50% of the values are shown in the shaded boxes, and the whiskers (the lines extending out from the box) extend out to the highest and lowest values excluding outliers[11].

[11] Outliers, if present, are represented by circles above or below the whiskers.

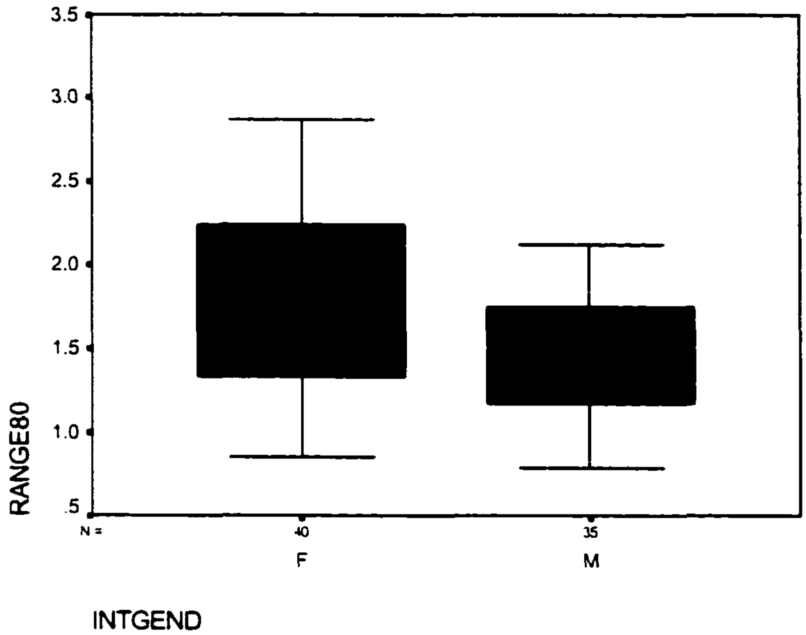


Figure 3. All 12 subjects, 80% pitch range by interlocutor gender, topics A and B (n = 40 with females, n = 35 with males).

Subjects also had larger standard deviations of pitch with female interlocutors, although the result did not reach statistical significance ($F(1, 73) = 2.801$, $p = 0.099$); means were 0.88752 ERBs (SD = 0.24612) with female interlocutors and 0.79917 ERBs (SD = 0.20550) with males, and medians were 0.885 ERBs and 0.773 ERBs.

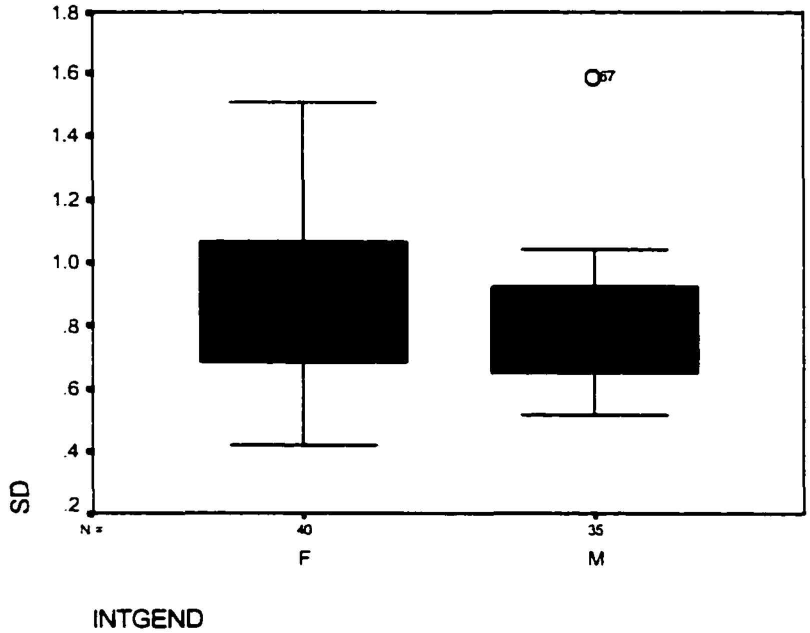


Figure 4. All 12 subjects, standard deviation of pitch by interlocutor gender, topics A and B (n = 40 with females, n = 35 with males).

Differences between median pitches according to the gender of the interlocutor were not significant, however. Means were 192.83 Hz (SD = 23.14) with female interlocutors and 188.4 (SD = 21.39) with males, and medians were 199.39 Hz and 195.11 Hz.

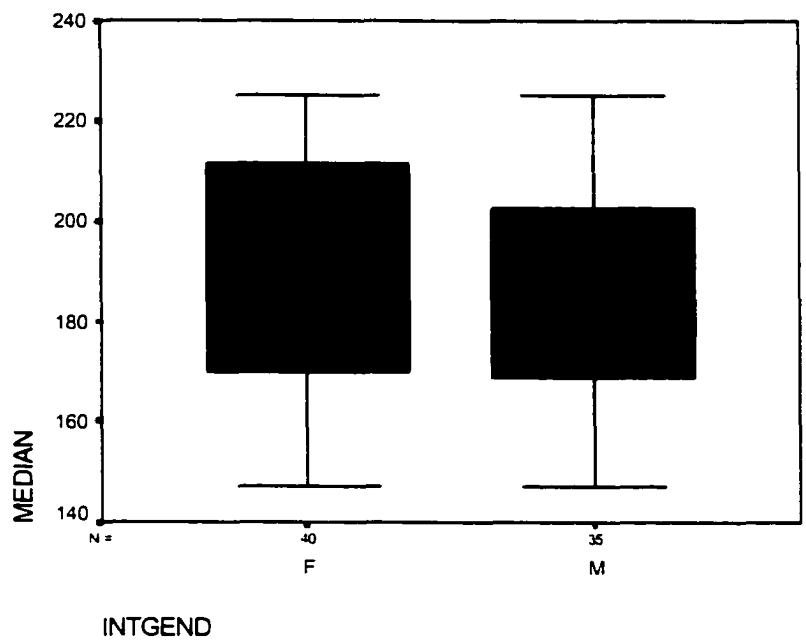


Figure 5. All 12 subjects, median pitch by interlocutor gender (n = 40 with females, n = 35 with males).

No comparisons based on the status of the interlocutor yielded significant results for the pooled data, as the three boxplots below show. The means and medians of the different measures for topics A and B were:

	MEDIAN PITCH (IN HZ)	80% PITCH RANGE (IN ERBS)	SD OF PITCH (IN ERBS)
Mean with peers	190.97	1.64	0.85
Mean with profs	190.53	1.54	0.84
Median with peers	198.16	1.59	0.805
Median with profs	195.11	1.42	0.806

Table 1. All 12 subjects, means and medians of the different measures for topics A and B.

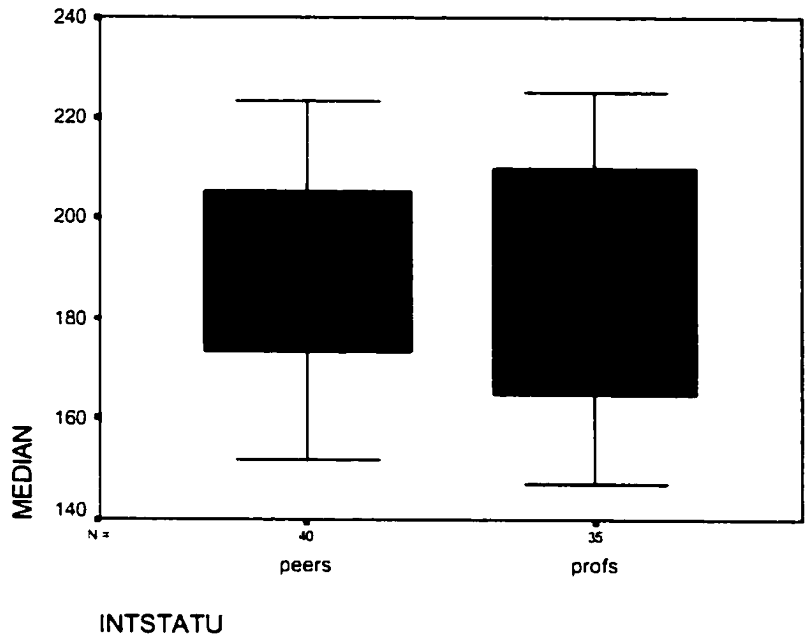


Figure 6. All 12 subjects, median pitch by interlocutor status (n = 40 with peers, n = 35 with professors).

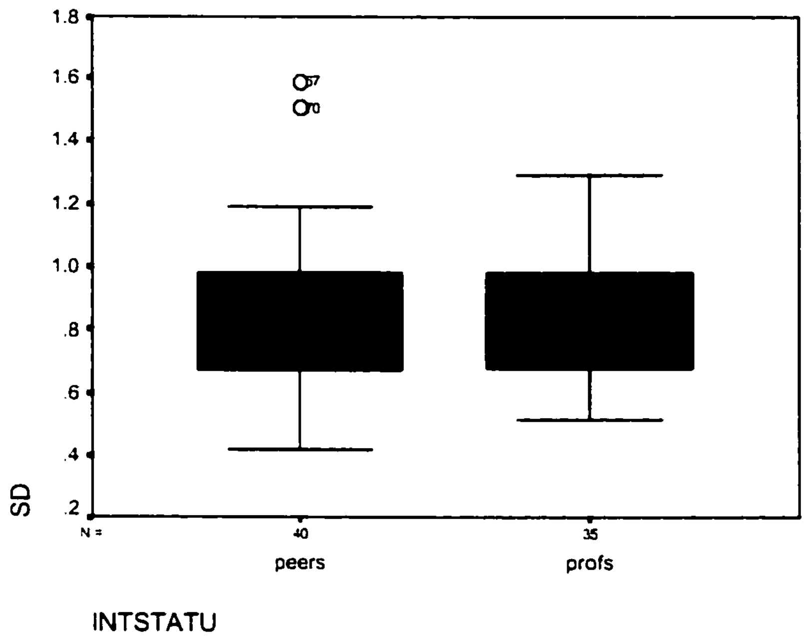


Figure 7. All 12 subjects, standard deviation of pitch by interlocutor status (n = 40 with peers, n = 35 with professors).

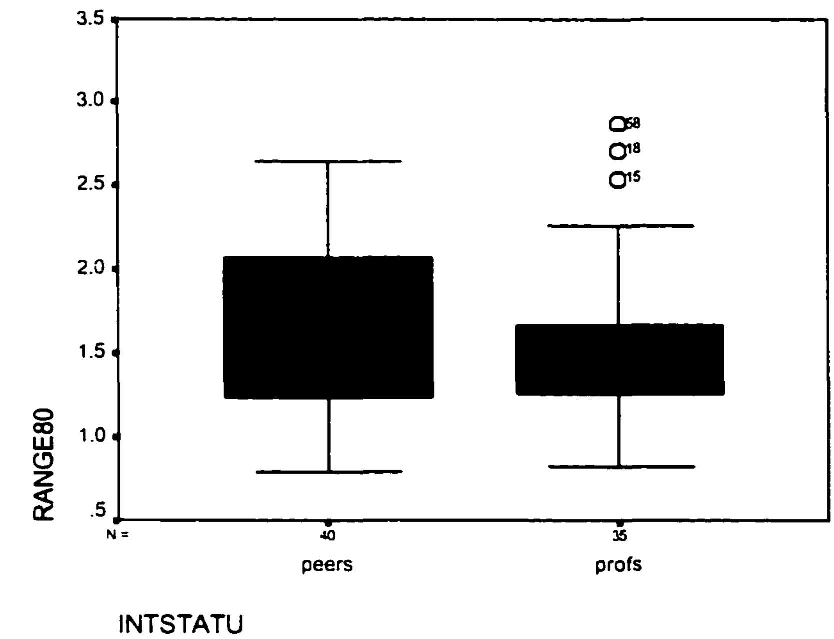


Figure 8. All 12 subjects, 80% pitch range by interlocutor status (n = 40 with peers, n = 35 with professors).

The hypothesis that subjects would use higher pitches in some manner with other females was upheld. While differences between median pitches and between standard deviations of pitch were not significant, pitch ranges were significantly higher, pointing to a wider range of pitches used with the female interlocutors.

For the data pooled from all 12 subjects, Univariate ANOVA revealed that there were no significant interactions between interlocutor gender and interlocutor status for any of the measures. However, there was a significant interaction between interlocutor gender and subject when subject was considered as a factor for pitch range ($F(11, 51) = 2.316$, $p < 0.05$); the behavior of individual subjects was involved in the significant effect of interlocutor gender on pitch range in the pooled data. (Subject also significantly interacted with interlocutor status for standard deviation of pitch and pitch range, but neither of these measures was significantly affected by interlocutor status considered alone in the pooled data.) These interactions indicate that, not surprisingly, different subjects behave differently. In order to explore the different behavioral trends within the subjects, I analyzed the results of each individual subject.

3.1.2. Individual Results and Subsequent Subgroups

The data from all 12 subjects analyzed together yielded some interesting results that need to be explained. However, it is likely to be the case that different individuals have different sociophonetic (here, pitch) strategies. I was concerned that alternative or additional patterns in the data among individual subjects were being masked by the patterns that emerged when all of the data were pooled. To explore whether or not pooling the data was obscuring sub-patterns in it, I analyzed the data of each individual subject. There is a statistical problem in analyzing and exploring every possibility within the data from each individual subject separately that needs to be acknowledged. Given that significance levels by definition indicate the likelihood that the finding is due to chance, there is a danger in running a large number of statistical tests on a single body of data in that it will be possible that the significant findings that are found will simply be due to chance. However, since individuals may vary in their behavior, it is worthwhile to look at their data separately, despite the statistical difficulties. Each different possible result for an individual was equally possible and would lead to a different interpretation of that individual's behavior, so it was necessary to test each possible factor for each subject. Although it is necessary to acknowledge that these results may in reality be due to chance, the significant results obtained via ANOVAs of pitch measures by interlocutor characteristics for each individual subject were:

	SIGNIFICANT RESULTS, $p < 0.05$	TRENDS, $p < 0.1$
S1	Median by gender, higher with females	
S2	Median by gender, higher with males SD by status, bigger with peers Range by status, bigger with peers	
S3	--	
S4	SD by gender, bigger with females Range by gender, bigger with females	
S5	SD by status, bigger with peers	Range by status, bigger with peers Median by status, higher with profs.
S6	Median by status, higher with profs.	
S7	Median by status, higher with profs.	SD by status, bigger with profs.
S8	Range by gender, bigger with females	SD by gender, bigger with females Median by gender, higher w/ females
S9	SD by gender, bigger with females	
S10	Median by gender, higher with females	
S11	--	
S12	Range by gender, bigger with females	SD by gender, bigger with females

Table 2. Significant differences in pitch measures according to interlocutor group for individuals.

When the number of data points is small, as it is within an individual subject's data, it may not be entirely appropriate to use of tests of significance. Tests of significance are intended to be used for data that are normally distributed. Large sets of data typically follow the normal curve, but smaller sets of data, such as the measurements for individual subjects, may not, and tests of significance may be less appropriate. Thus, the results of these tests may or may not tell the full story in and of themselves. However, I am using them as indicators as to which subgroup a given subject might belong to in the subgrouping process described below. For the pooled data above and the individual and pooled subgroup data below, I provide both means and medians of the data to give the most detail and the clearest picture of the results of the comparisons being made, regardless of the normality of the data distributions.

3.1.2.1. The Gender Group

Looking at all of the results for individual subjects revealed some interesting findings; there appear to be two different trends in the data, the first of which will be described in this section. Several subjects had significant findings based on the gender of the interlocutor but not any based on the status of the interlocutor (S1, S4, S8, S9, S10, and S12). The means and standard deviations and medians of the significant results (at the 0.05 level) for these individual gender-only subjects are listed below (all significant findings here were in the topic A and topic B data, pooled together, except for the one noted). Median pitch is reported in Hertz and standard deviations of pitch and pitch ranges are reported in ERBs.

	MEASURE		MEAN (SD)	MEDIAN
S1	Median	w/ F, n = 3	216.92 (6.92)	213.23
		w/ M, n = 3	199.52 (1.82)	199.62
S4	SD	w/ F, n = 2	0.89 (0.09)	0.89
		w/ M, n = 3	0.68 (0.06)	0.70
	Range	w/ F, n = 2	1.35 (0.01)	1.35
		w/ M, n = 3	0.96 (0.15)	1.04
S8	Range	w/ F, n = 4	2.5 (0.16)	2.47
		w/ M, n = 3	1.95 (0.11)	1.97
S9	SD	w/ F, n = 3	1.08 (0.1)	1.12
		w/ M, n = 3	0.89 (0.03)	0.90
S10	Median	w/ F, n = 2*	231.75 (0.41)	231.75
		w/ M, n = 2	223.17 (0.2)	223.17
S12	Range	w/ F, n = 3	2.35 (0.48)	2.27
		w/ M, n = 3	1.1 (0.06)	1.11

*Result from analysis of first minute of interviews.

Table 3. Means and medians of significant results for gender-only subjects.

What was striking about this "gender-only" group of subjects, however, was that within the group, the findings were all in the same direction; these subjects all had higher median pitches, bigger pitch ranges and/or bigger standard deviations of pitch with female interlocutors than they did with male interlocutors. When the data from these six subjects are pooled together, the results of ANOVAs of pitch range and SD of pitch by interlocutor gender are significant, and at levels lower than any in any individual subject's data ($F(1, 35) = 11.633$, $p < 0.01$, and $F(1, 35) = 5.536$, $p < 0.05$). While it may appear to be obvious that pooling subjects with the same trend in their data would lead to a stronger trend, it is not in fact a certainty. Thus, I will report the results of ANOVAs for all pooled subgroups, as well as their means and medians for the relevant measures. For standard deviation of pitch for the gender-only group, means were 0.91 ERBs (SD = 0.23) with female interlocutors and 0.76 ERBs (SD = 0.15) with males, and medians were 0.95 ERBs and 0.71 ERBs.

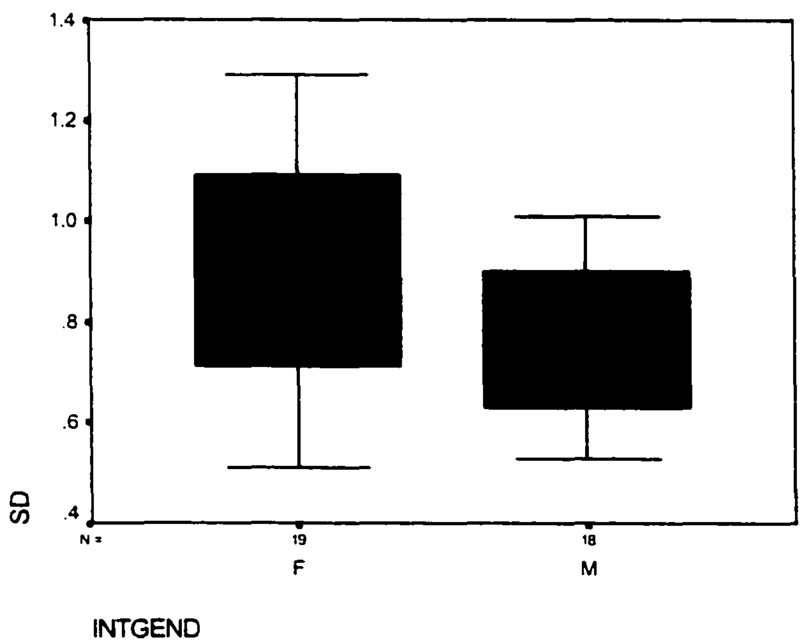


Figure 9. Gender group (S1, S4, S8, S9, S10, and S12), standard deviation of pitch by interlocutor gender, topics A and B (n = 19 with females, n = 18 with males).

For 80% pitch range, means were 1.86 ERBs (SD = 0.56) with females and 1.33 ERBs (SD = 0.36) with males, and medians were 1.78 ERBs and 1.22 ERBs.

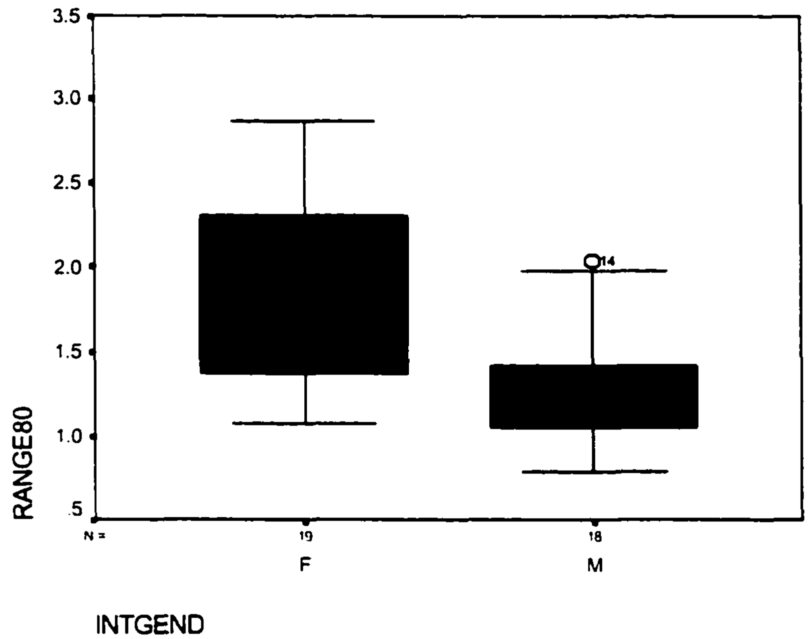


Figure 10. Gender group (S1, S4, S8, S9, S10, and S12), 80% pitch range by interlocutor gender, topics A and B (n = 19 with females, n = 18 with males).

While median pitch by interlocutor gender was not statistically significant for the pooled gender group data, the medians were higher with female interlocutors (but there is a large spread in the data). Means were 197.65 Hz (SD = 23.25) with females and 188.03 Hz (SD = 23.84) with males, and medians were 204.34 Hz and 192.97 Hz.

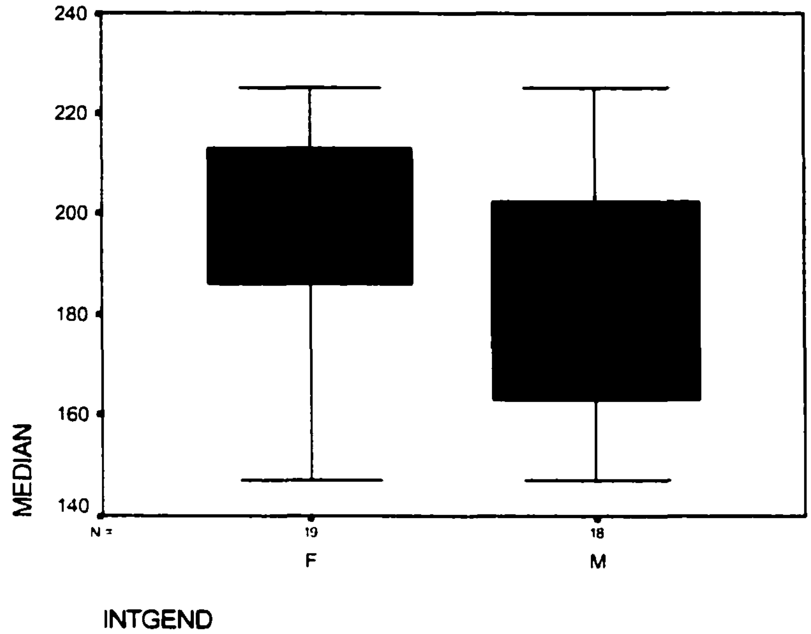


Figure 11. Gender group (S1, S4, S8, S9, S10, and S12), median pitch by interlocutor gender, topics A and B (n = 19 with females, n = 18 with males).

It seems that if a subject was affected only by the interlocutors' gender and not their status, she used higher and/or more varied pitches with women. This kind of pitch pattern may point to a high-engagement style or camaraderie politeness strategy that may occur in female-female conversations (see Chapter 4).

The only other subject who had a significant result based on interlocutor gender was S2, who also had significant results based on interlocutor status. She had higher median pitches with males than she did with females (means were 165.1 Hz (SD = 5.43) with females and 175.6 Hz (SD = 2.91) with males, and medians were 164.44 Hz and 174.41 Hz with females and males, respectively). Her data, the fact that four subjects in the gender-only group had significant SD and/or range results but only two had significant median results, and the fact that median results in the pooled gender group data were not significant caused me to wonder whether or not it was subjects' standard deviations of pitch and pitch ranges that are the crucial elements in this subgroup's pitch patterns, rather than the median pitch. Standard deviation of pitch and pitch range positively correlate across the data, i.e., when one gets larger/smaller, so does the other. They may in fact index one larger (e.g., pitch variability) aspect of pitch usage. The idea of this subgroup's data and underlying sociophonetic strategy hinging on pitch variability is consistent with the explanation that this pitch pattern, with its use of many different pitches and frequent pitch changes, is part of an animated, high-engagement conversational style. It may be that S1 and S10 are not part of this gender-based group but form their own gender group with a different sociophonetic or politeness strategy. Another possible explanation, explored in Experiment 2, is that their pitch patterns are due to audience accommodation rather than some kind of audience expectation-based conversational strategy. However, it may be that they are part of this gender group and their higher median pitches are a slightly different way of sociophonetically communicating camaraderie with their female interlocutors.

3.1.2.2. The Status Groups

Four subjects had significant findings at the 0.05 level based on the status of the interlocutor (S2, S5, S6, and S7). The means and standard deviations and medians of the significant results for these individual subjects are listed below (all significant findings here were in the topic A and topic B data, pooled together). Median pitch is reported in Hertz and standard deviations of pitch and pitch ranges are reported in ERBs.

	Measure		Mean (SD)	Median
S2	SD	w/ peers, n = 4	1.04 (0.10)	1.01
		w/ profs., n = 3	0.85 (0.08)	0.81
	Range	w/ peers, n = 4	2.27 (0.31)	2.27
		w/ profs., n = 3	1.63 (0.18)	1.72
S5	SD	w/ peers, n = 3	1.43 (0.21)	1.51
		w/ profs., n = 4	1.06 (0.08)	1.06
S6	Median	w/ peers, n = 4	198.52 (4.04)	197.98
		w/ profs., n = 3	209.84 (6.37)	210.04
S7	Median	w/ peers, n = 2	186.23 (0.59)	186.23
		w/ profs., n = 2	196.09 (2.54)	196.09

Table 4. Means and medians for significant results for status-based subjects.

The results of these subjects do not point to one cohesive strategy to the extent that the gender-only subjects' do, but there are suggestive patterns within them. While S2, as previously mentioned, had one significant result based on interlocutor gender, the other three had none. However, it does not seem to be the case that the three status-only subjects behave similarly. Instead, S6 and S7 have the same pattern in that they both have higher median pitches with professors versus with peers, and S2 and S5 pattern together in that they both have bigger standard deviations of pitch and/or pitch ranges with peers. The sub-patterning between S6 and S7 and between S2 and S5 lead me to suspect that there are two status subgroups, each with its own sociophonetic (pitch) strategy. Status group 1 (S6 and S7) could be raising the general level of pitches used (median pitch) as a deference politeness strategy while the larger pitch variability (SD and range) of status group 2 (S2 and S5) could reflect the usage of a camaraderie politeness style or a higher level of comfort with peers.

The pooled data for these suggested subgroupings also lead to some significant results, but, again, they are not as striking as those for the gender-only group. When S6's and S7's data are pooled, the results of an ANOVA of median pitch by interlocutor status is not quite significant ($p = 0.069$), but the means (194.42 Hz ($SD = 7.08$) with peers and

204.34 Hz (SD = 8.87) with professors) and medians (195.75 Hz with peers, 203.37 Hz. with professors) are suggestive.

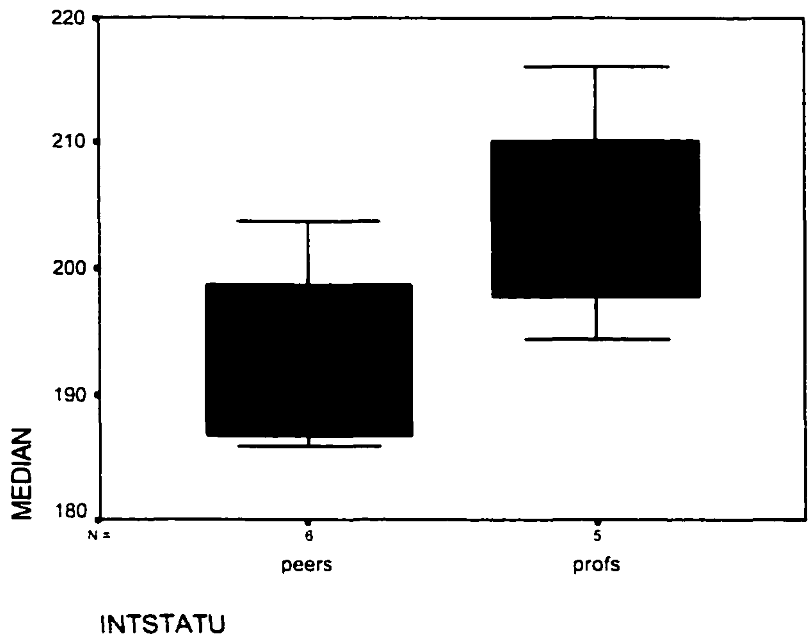


Figure 12. Status group 1 (S6 and S7). median pitch by interlocutor status, topics A and B (n = 6 with peers, n = 5 with professors).

The standard deviations of pitch and pitch ranges were not significantly different for status group 1, but they were slightly larger with professors. Mean standard deviation of pitch was 0.62 ERBs (SD = 0.10) with peers and 0.68 ERBs (SD = 0.10) with professors, and median SD's were 0.64 and 0.70. Mean pitch range was 1.22 ERBs (SD = 0.25) with peers and 1.34 ERBs (SD = 0.29) with professors, and median ranges were 1.26 and 1.49.

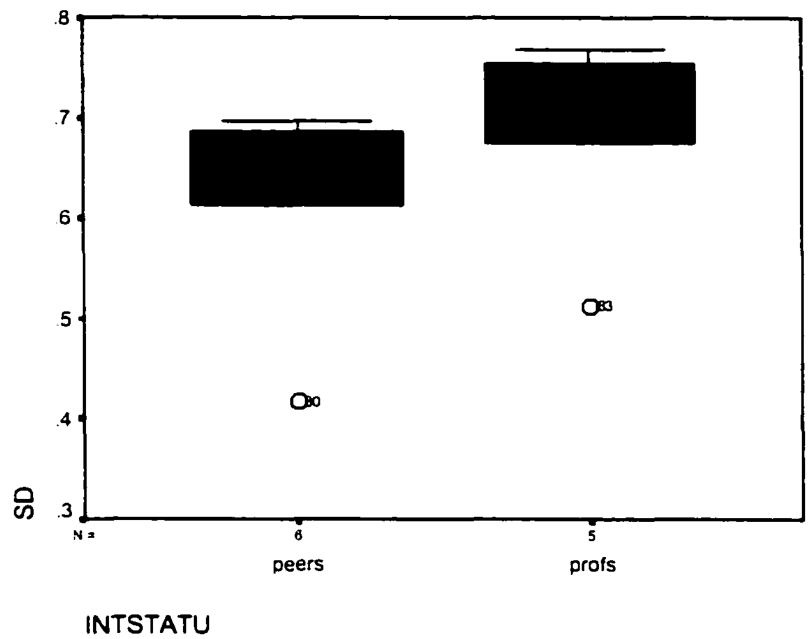


Figure 13. Status group 1 (S6 and S7), standard deviation of pitch by interlocutor status, topics A and B (n = 6 with peers, n = 5 with professors).

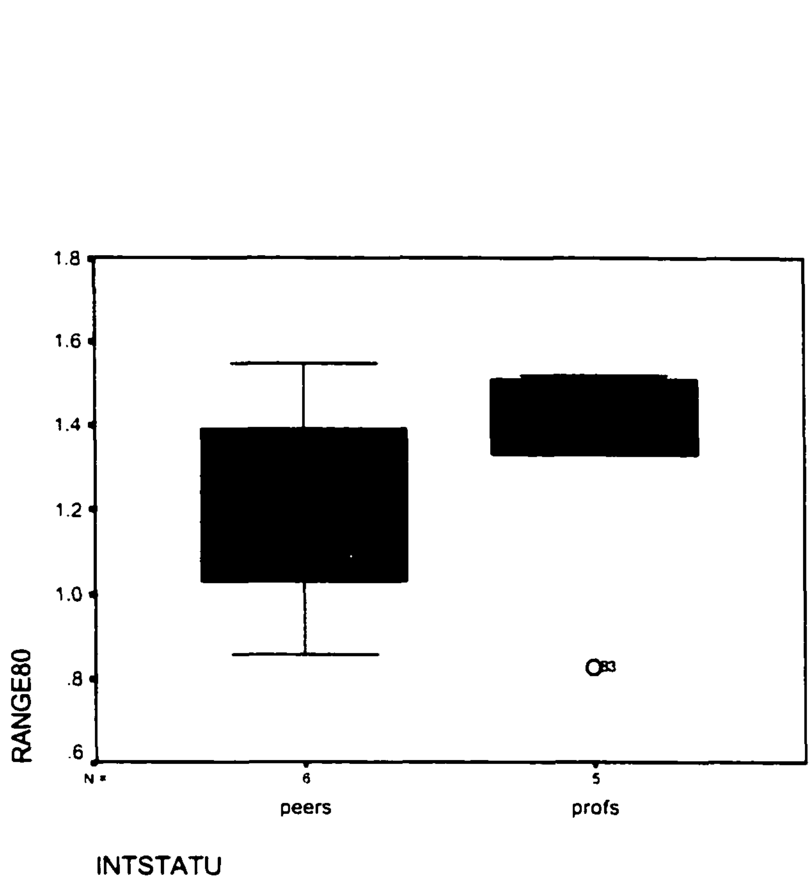


Figure 14. Status group 1 (S6 and S7), 80% pitch range by interlocutor status, topics A and B (n = 6 with peers, n = 5 with professors).

When S2's and S5's data are pooled, two significant results are found, for pitch range ($F(1, 12) = 16.707$, $p < 0.01$) and for SD of pitch ($F(1, 12) = 4.89$, $p < 0.05$). The means for pitch range were 2.17 ERBs (SD = 0.40) with peers and 1.47 ERBs (SD = 0.20) with professors, and the range medians were 2.12 and 1.42. The means for standard deviation of pitch were 1.21 ERBs (SD = 0.25) with peers and 0.97 ERBs (SD = 0.13) with professors, and SD medians were 1.18 and 0.97. Status group 2 median pitches were not significantly different according to the status of the interlocutor. Means were 165.27 Hz (SD = 10.33) with peers and 163.19 Hz (SD = 7.91) with professors, and medians were 172.12 and 161.9.

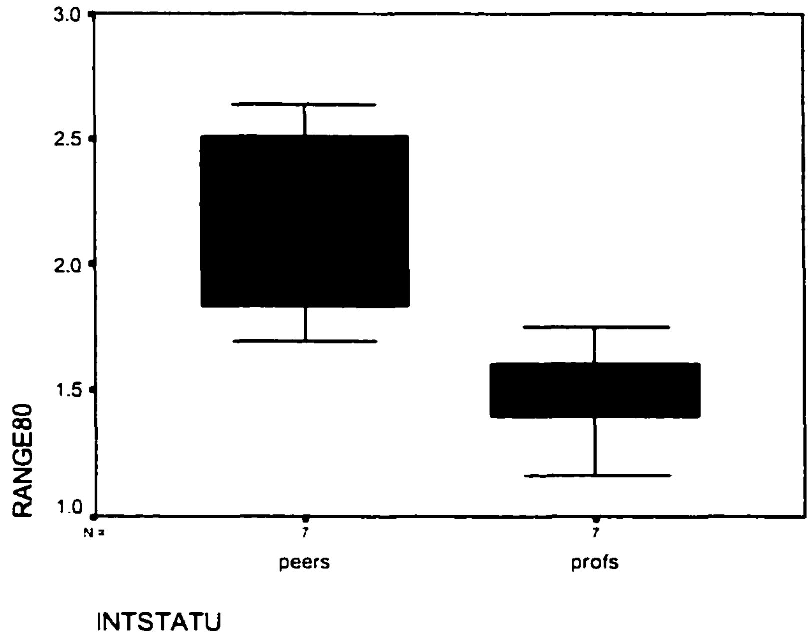


Figure 15. Status group 2 (S2 and S5), 80% pitch range by interlocutor status (n = 7 with peers, n = 7 with professors).

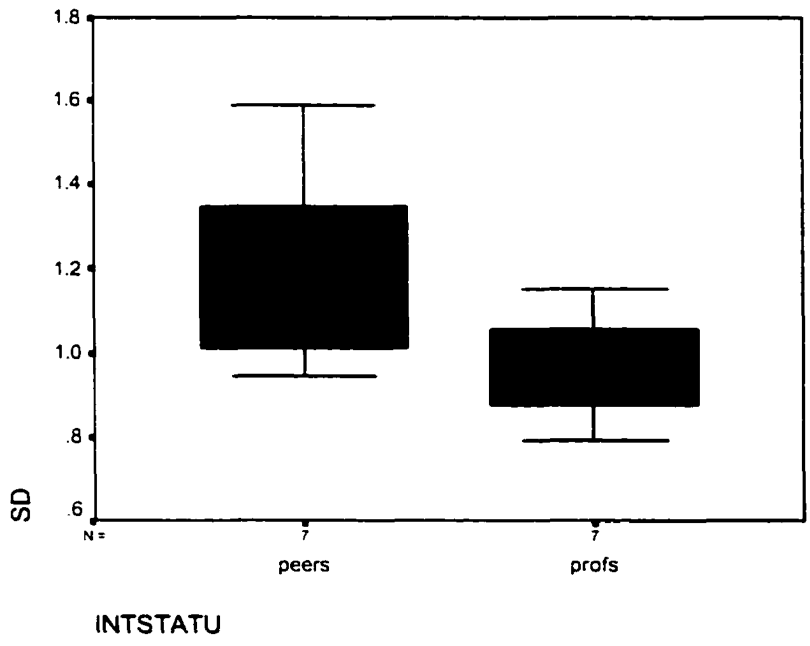


Figure 16. Status group 2 (S2 and S5), standard deviation of pitch by interlocutor status (n = 7 with peers, n = 7 with professors).

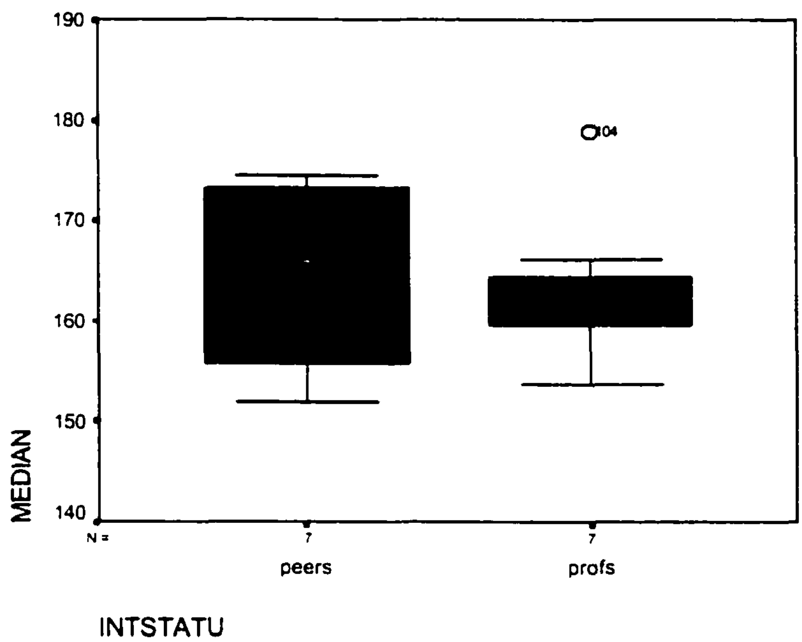


Figure 17. Status group 2 (S2 and S5), median pitch by interlocutor status (n = 7 with peers, n = 7 with professors).

At first glance, the results of the three measures for the two status subgroups look contradictory. It appears that group 1 uses higher and more varied pitches with professors while group 2 does so with peers. However, the strongest result for group 1 was for median pitch (as in the data of the individuals involved, as expected), and for status group 2, the strongest results were for SD of pitch and pitch range. The strongest results point to what is key in the strategies of each group; group 1 raises the pitch tessatura in deference to professors (and pitch variability is perhaps affected secondarily), and group 2 uses heightened pitch variability with peers (and median pitch is perhaps affected secondarily) as part of a camaraderie politeness style. Although the two status subgroups are using different strategies that lead to separate pitch patterns, there is a consistency between them in that they both can be seen as reflecting more formality and deference or less comfort with the professors than with the peers, which makes sense, intuitively (see Chapter 4 for more discussion).

This theoretical water is slightly muddied if one considers the results at the level between 0.05 and 0.1, which I have deemed trends. The results in this category for status are S5 having a higher median with professors and a bigger pitch range with peers. However, the former finding is in the first minutes of the interviews, while the latter, along with the significant SD finding for S5, are from topics A and B, so the samples are overlapping but different. If these results are considered, S5 seemingly straddles the two subgroups, perhaps using both deference and camaraderie strategies where appropriate. The other subject with a finding at this level of significance is S7, who had a bigger SD of pitch with professors. This finding may contradict the idea that more variability points to a camaraderie politeness style. It is difficult, however, to know how seriously to consider these findings, as, statistically, they are more likely to be due to chance. (For the gender-only group, the findings at this significance level, i.e., p is between 0.05 and 0.1, are consistent with the significant, i.e., $p < 0.05$, results.)

Regardless of the interpretations of the strategies used by the different subject subgroups, there is some independent evidence for the make-up of the subgroups themselves. Using the hierarchical clustering function in SPSS, the following clusters of subjects were found using the first minutes of pitch data for all three measures from the interviews of all 12 subjects with all four interlocutors:

A: 1, 6, 7, 10, 11
B: 2, 5
C: 3, 8, 9
D: 4, 12

The dendrogram is shown in Figure 18.

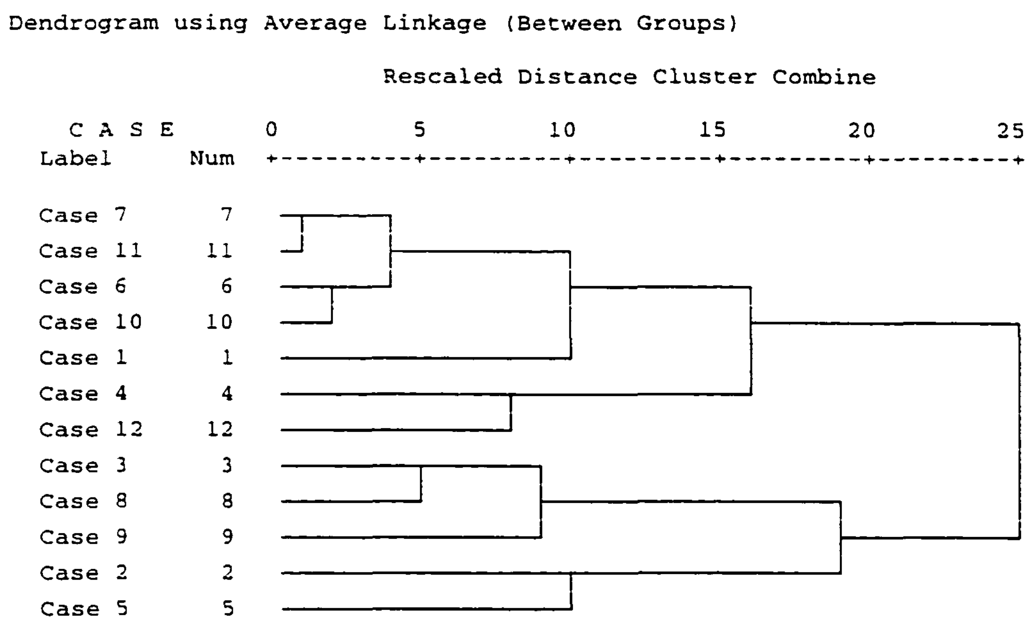


Figure 18. Dendrogram for all 12 subjects (cases), first minutes of interviews.

When I tried to apply the same function using the pitch measurements from excerpts from topics A and B in the interviews, however, subjects with missing values (i.e., if any of the subject's excerpts were less than 40 solid seconds and were not used in experiment one), were excluded by the SPSS hierarchical clustering function. If all topic A and B excerpts were considered, only one subject had a full range of data. To combat this problem, I analyzed parts of the data (i.e., topic A with the male peer, topic B with the female professor, and topic A with both female interlocutors), and only used subjects with complete data. This analysis created the following clusters, which do not include subjects 1, 3, 4, and 7, as they had missing data:

A: 2, 5
B: 6, 10, 11
C: 8, 9
D: 12

The dendrogram is shown in Figure 19.

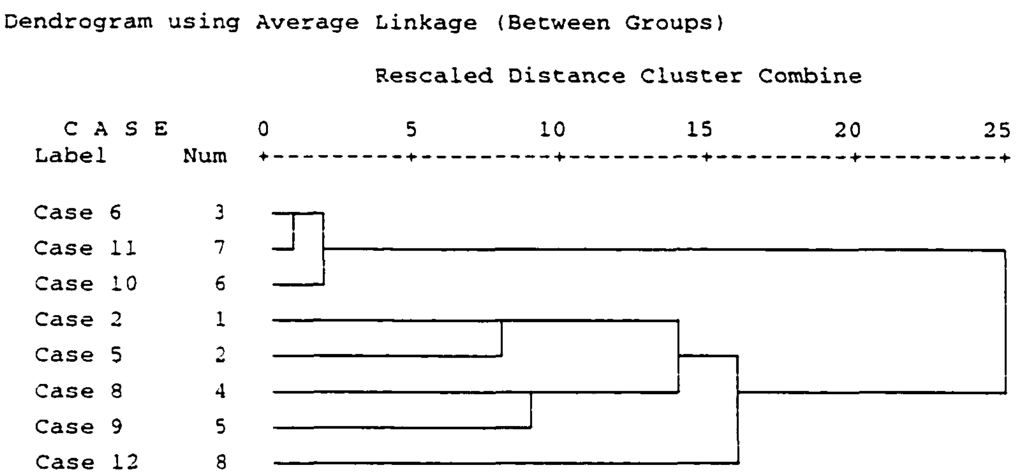


Figure 19. Dendrogram for subjects 2, 5, 6, 8, 9, 10, 11, and 12, A topic a's, B topic a's, P topic a's, and P topic b's.

The two kinds of data (first minutes and topics A and B) produce consistent clusters, clusters that provide secondary evidence for the clusters (subgroups) that I formed based on the significant results of individual subjects. Hierarchical clustering provides evidence for S2 and S5, S6 and S7, S8 and S9, S4 and S12, and S1 and S10 clustering together, respectively. While the clusters group subgroups together slightly differently (the gender group subjects are not all grouped together, and S1 and S10 are grouped together with S6 and S7), the results of the hierarchical clustering are remarkably consistent with those based on the statistical patterning reported above.

3.1.2.3. Results Based on Topic Differences

There were no significant results at the 0.05 level for any pitch measure based on the type of topic under discussion for all 12 subjects pooled together or for any individual subject. One subject (S9) had one trend; her median pitch was higher for topic B (mundane) segments than for topic A (gendered) segments ($F(1, 4) = 5.768$, $p < 0.1$). Means were 198.98 Hz (SD = 5.32) for topic A and 210.16 Hz (SD = 5.52) for topic B. Possible explanations for the lack of significant findings based on topic type will be explored in Chapter 4.

3.1.2.4. The Questionnaires

The Personal Attributes Questionnaire was scored according to the instructions given in Spence and Helmreich (1978). Responses to questions were recorded and totaled so that a score for each scale (masculinity, femininity and masculinity-femininity) was obtained for each subject. The Feminist Perspectives Scale was scored according to instructions provided with the survey questions (N. Henley, personal communication). Responses to the liberal, radical, and cultural feminist subscales were recorded and combined into a "Femscore," a summary feminism score, for each subject. Conservative scale scores were also separately obtained.

Using SPSS, I ran ANOVAs to explore the possibility that subjects' survey responses could be responsible for their different pitch patterns according to their interlocutor, i.e., to test the hypothesis that subjects who had higher femininity, masculinity, feminism, or conservatism scores might behave differently from those who had lower scores. It may be the case that the subgroups of subjects described above (the gender group and the status groups) had responded differently to one or both surveys, and these personality differences correlated with their different pitch behavior patterns. This was not the case. The means and standard deviations for each survey scale for the three subgroups follow. The gender group consists of four subjects, and each status group consists of two subjects.

	GENDER GROUP	STATUS GROUP 1	STATUS GROUP 2
Masculinity	23.0(4.08)	21.5(0.71)	22.5(0.71)
Femininity	24.75(2.22)	25.5(2.12)	22(0)
Masc.-Fem.	17.0(3.92)	12.5(3.54)	16.5(3.54)
Feminism	51.75(15.82)	49.5(6.36)	51.5(4.95)
Conservatism	11.25(4.35)	6(0)	12.0(7.07)

Table 5. Means for each survey scale for the three subgroups. Masculinity, femininity, and masculinity-femininity scores are each out of a possible 32. Femscore (feminism score) is out of a possible 105 and Conservatism is out of a possible 35.

None of these differences between subgroups is significant. There are few patterns to observe. Status group 1 is slightly less masculine, more feminine, and less conservative, but again, these differences are not statistically significant. If S1 and S10 are included in the gender group, the results are still not significant.

There are many problems with the use of such questionnaires. Aside from the difficulty of designing them, it is difficult to know beforehand what personality attributes or attitudes might be responsible for a particular behavior pattern. Furthermore, people may consciously or quite possibly unconsciously not be completely honest; people may not be completely self-aware. What is truly needed to evaluate if or how personal identity characteristics are influencing the politeness strategies and resultant pitch patterns that speakers choose is an ethnographic study within a speech community. This kind of study will be an important future step towards uncovering and understanding motivations behind speakers' strategies.

3.2. Experiment Two: Exploration of Pitch Accommodation between Interlocutors

Another possible explanation for some of the results found in experiment one that needs to be explored (or ruled out) is addressee accommodation. As discussed in Chapter 1, audience design can involve converging towards the speech characteristics of an interlocutor, i.e., accommodating to the audience (Bell, 1984). It may be that such accommodation happens in the realms of pitch tessitura or pitch variability, as well as in those of vowel quality, lexical choice, etc. It is necessary to discern whether or not the results of experiment one in general or in part are due to audience accommodation rather than differential use of politeness strategies based on the gender or status of the interlocutor. I was in particular curious about the results from subjects who diverged in some manner from the pattern that other subjects' data pointed to. For example, S2 was the only subject to have significant results based on both gender and status of the interlocutor, and was also the only subject to have higher median pitch with males, in contrast with the other subjects who had significant results based on interlocutor gender who had higher or more variable pitches with females. Among these six subjects, four (S4, S8, S9, and S12) had more variable pitches with females and two (S1 and S10) had higher median pitches with females. S1 and S10, with significant results for median pitch rather than standard deviation of pitch or pitch range (i.e., pitch variability), can also be considered a minority pattern that needs explanation.

Aside from testing whether pitch accommodation is responsible for any results from experiment one, it is a possibility that needs to be explored in and of itself. Firstly, does it happen? If it happens, how does it happen? Who does it and to whom? Does the gender or the status of the interlocutor affect whether or not an interlocutor is accommodated to? The data as whole needs to be examined for patterns.

Before turning to the discussion of the accommodation results relevant to experiment one, it is necessary to sort out the different possible explanations for certain pitch patterns. This project explores and tries to differentiate between two kinds of social adjustments of pitch. The first one, explored in the first experiment, deals with how (female) speakers might adjust their pitches according to whom their interlocutor is, e.g., their gender, their status, etc. They might speak according a preconceived notion about the appropriate way to speak to a given individual. The second kind of social pitch adjustment is that of accommodation, explored in the second experiment. It deals with how speakers might adjust their pitch in some fashion according the pitches used by their interlocutor. The crucial distinction between the two kinds of social pitch adjustments is that in accommodation, the speaker adjusts his or her pitch according to what he or she hears. *Accommodation is a response to what is heard*, unlike expectation-based adjustments (expectations about the interlocutor and what is thus situationally appropriate). They are separate phenomena, but distinguishing between the two is not always an easy task. The case for accommodation is relatively clear when the speakers adjust their pitches throughout a conversation and their pitch patterns converge. As discussed below, my working hypothesis is that this type of convergence will be reflected in a correlation between the two speakers' pitch measurements over the course of an interview. However, one situation I have encountered in the data where the line between accommodation and expectation is unclear is where there is not a lot of variation in either speakers' measurements over time. Without variation in the data, there will not be significant correlation, as correlation measures whether sets of numbers co-vary. If a given speaker and interlocutor both have normalized scores for a pitch measure in a certain small range that is set apart from the rest of that speaker's data, does it mean there is a kind of global accommodation where the speakers are convergent in their general level of median pitch or general level of pitch variability, without varying them over time in a similar way? Or is it a case of convergent expectation-based pitch adjustment? I do not believe that every case can be attributed to one or the other explanation without any doubt. With this understanding in mind, I have examined the data relevant to experiment one, and then the data as a whole, each of which is reported in the following two sections.

3.2.1. On Explaining Experiment One Results

Looking at the results from experiment two to explain those from experiment one is slightly problematic because the data used for the pitch measurements in each experiment were somewhat different. In experiment one, the first minute of solid speech for two kinds of topics of conversation and for the chatting at the beginning of the interview were measured. The data were prepared by hand and pauses, interlocutor speech, overlapping speech and laughter were all edited out. In experiment two, computer scripts prepared the data from throughout the interviews so that pauses, interlocutor speech, and overlapping speech were not measured, but some laughter was perhaps included in what was measured, although much of it occurred during overlapping speech portions. However, despite these differences, it is still useful to test whether the measurements from the second experiment can explain those from the first. Generalizations have been drawn from both sets of data, and those from the second can provide at least the first piece of evidence as to whether the first ones have pitch accommodation at their root.

The same statistical problem reported above for experiment one is present in experiment two, as well. Because it is entirely possible that some subjects accommodate while others do not, it was necessary to measure accommodation for individual subjects (i.e., for each individual with different interlocutor groups and within individual interviews), rather than pooling them together. Given that significance levels by definition indicate the likelihood that a finding is due to chance, again, there is a danger in running a large number of statistical tests on a single body of data in that it will be possible that the significant findings that are found will simply be due to chance. However, since individuals may vary in their behavior, it is necessary to look at their data separately, despite the statistical difficulties. Thus, correlations were measured for the three pitch measures for individuals, and these are the results that will be reported as they are relevant to experiment one and in exploring trends present in the data as a whole.

In order to determine whether pitch accommodation was behind the results from experiment one, all of the interviews involved in the significant results for experiment one for each subject were checked for evidence of accommodation. The test used was a bivariate, two-tailed Pearson correlation within SPSS. The Pearson correlation is a measure of linear correlation between two sets of data. A simple correlation could reflect accommodation since correlation measures concomitant variation between two sets of numbers. Grossly stated, sets of pairs of numbers correlate when one member of a pair can to some degree be used to predict the other; the two sets of numbers are related to each other. When the two sets of numbers are plotted as points on an x-y axis, i.e., as

Cartesian coordinates, and they correlate, the linear relationship will be evident. If a regression line were drawn, the points would cluster around it, and the slope of that line would reflect the nature of the relationship. As an example of how I used correlation measures to check experiment one results for evidence of accommodation, I will discuss S6. Since S6 had a significantly higher median pitch with professors than she did with peers, I measured the level of correlation between S6's median pitches and her interlocutors' median pitches for her interviews with peers as a group and for her interviews with professors as a group, as well as for each individual interview. I did this by running a Pearson correlation for S6's normalized scores (i.e., weighted averages normalized using S6's median of median pitches throughout all of her data, see Chapter 2) and her interlocutors' normalized scores for median pitches. The correlation measurements were based on a normalized score for median pitch for S6 and for the interlocutor for each of the overlapping minute-long time chunks in each of her four interviews. Since the rate of change measure (i.e., the delta measure, the difference between the normalized scores for consecutive time chunks) wasn't measured in experiment one, correlations involving it are less relevant for determining accommodation's role, if any, in experiment one's results, but they were also measured and will be reported.

If pitch accommodation is responsible for the higher median pitch with professors for S6, there should be evidence of accommodation in one or both of the relevant data sets, i.e., S6 with the peer interlocutors and S6 with the professor interlocutors. The amount of correlation did not reach the level of significance for either group of interviews for either normalized scores or deltas. There were no significant correlations for any of the individual interviews for median pitch, either. Another subject who had two significant results in experiment one based on interlocutor status, S2, who had a larger SD and range with peers, had no significant correlations for these pitch measures for the professors as a group or in either individual interview (recall that interviews with the peer interlocutors were unavailable due to poor sound quality in the interlocutors' channel).

There were two other subjects who had experiment one results based on interlocutor status, S7, who had higher median pitches with professors, and S5, who had a larger SD of pitch with peers. They did have some significant findings in the relevant interviews/measures. S7 had a significant positive correlation for normalized scores of median pitches when all four interviews were grouped together, $r = 0.209$, $p < 0.05$, $n = 100$.

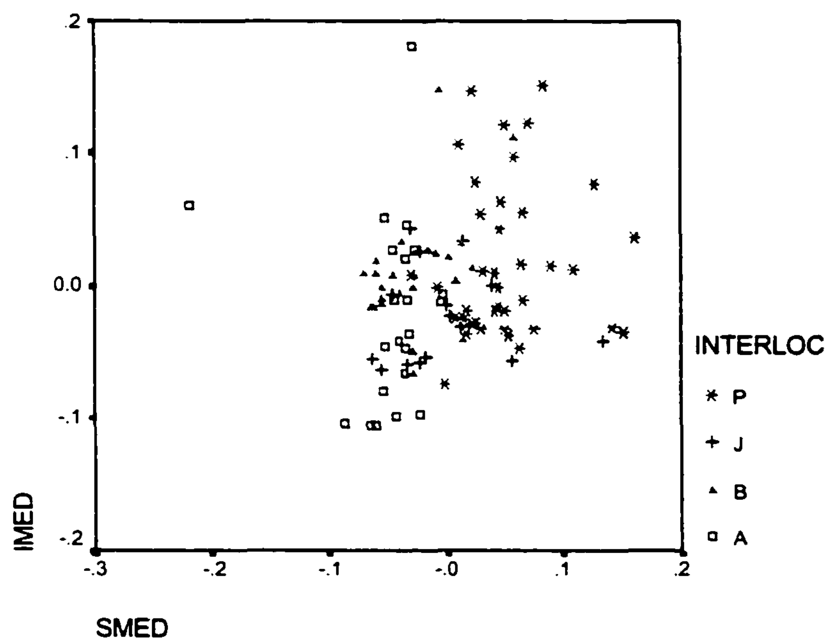
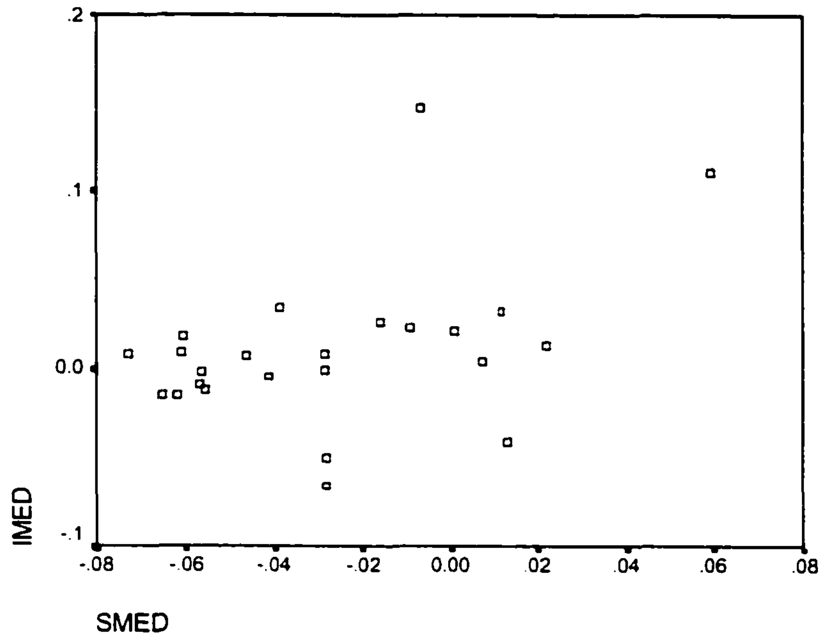


Figure 20. S7 with all four interlocutors, normalized scores, medians.

However, as will be discussed in the following section on overall accommodation trends, correlations within groups of interviews can reflect many things. This group scatterplot suggests the situation discussed above where it may be impossible to distinguish between a kind of global accommodation in pitch (not involving co-variation) and a kind of expectation-based social adjustment based on the perception of the interlocutor in that S7 and P's data points are both more extreme that the rest of the data. (Throughout this chapter, individual interlocutors will be referred to by their first initial; A = female peer, B = male peer, P = female professor, and J = male professor.) However, unlike in S4's range data described below, the normalized scores for the entire interviews do not support the accommodation interpretation in that while S7 has her highest entire interview-based median pitch with P, P does not have her highest entire interview score with S7, but rather has her fourth lowest score with her. (See Appendix 3 for normalized scores calculated for the interviews as wholes for all subjects and interlocutors.)

Another factor commonly involved in explaining the pattern in a group of interviews and involved here is the effect of one or more individual interviews with a correlation within the group. In this case, the individual interview was that between S7 and B, the male peer. Their median pitches positively correlated, $r = 0.416$, $p < 0.05$, $n = 24$. For significant correlations within individual interviews, line graphs will be presented to show the data points as they occurred across time, in addition to the scatterplots to show the correlation.



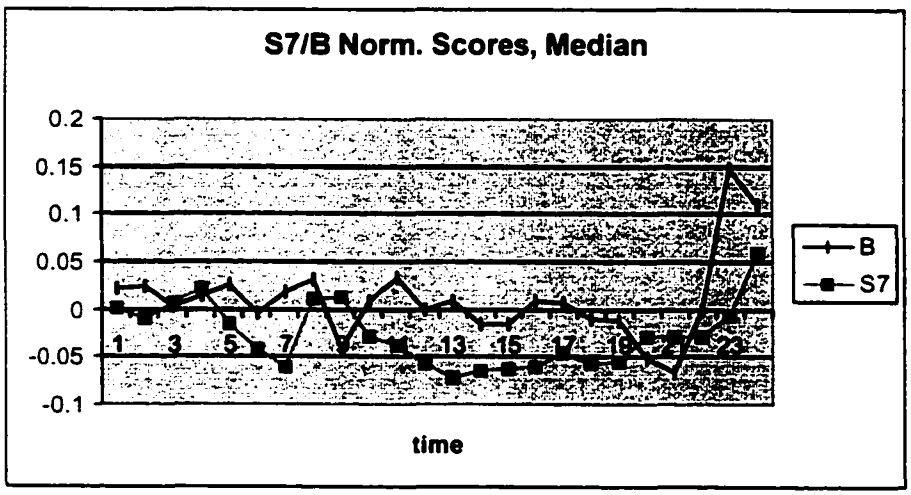


Figure 21. S7 with B, normalized scores, medians.

S7 did indeed have a significant correlation with B, but she did not with A, the female peer, or with P or J, the professors.

S5 had no significant correlation for normalized scores for the relevant measure, standard deviation of pitch. However, she did have a group of correlations involving the delta measure for SD; she had a significant positive correlation when all four interviews were grouped together and when just the two peer interlocutors were grouped together.

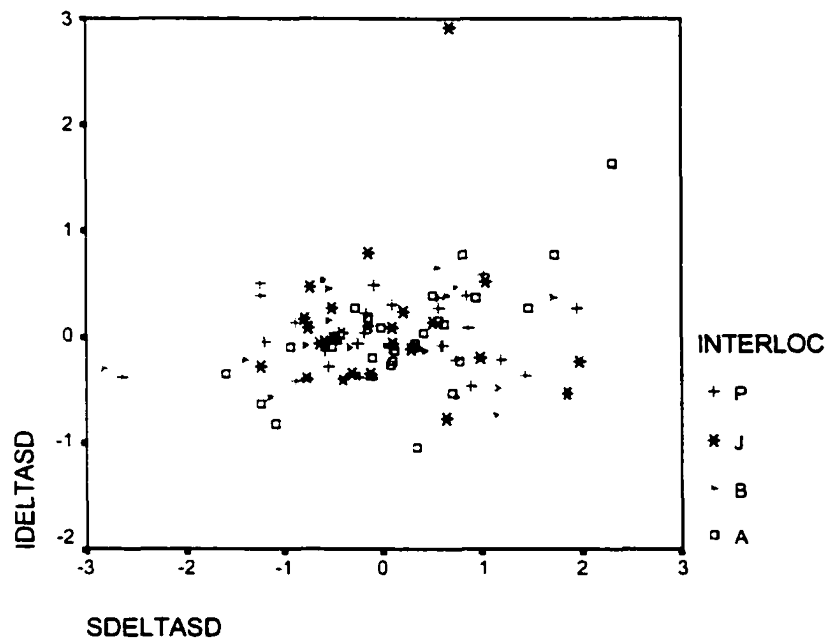


Figure 22. S5 with all four interlocutors, deltas, SDs.

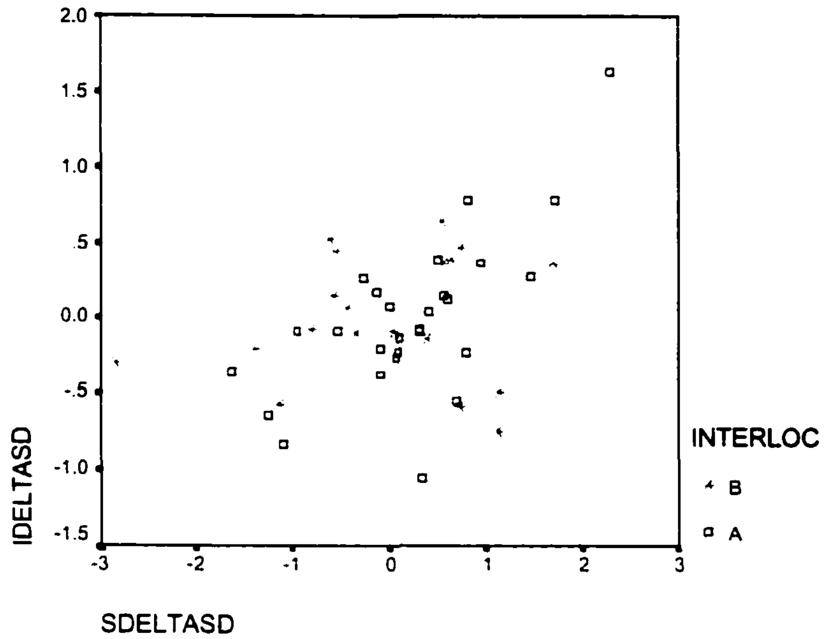
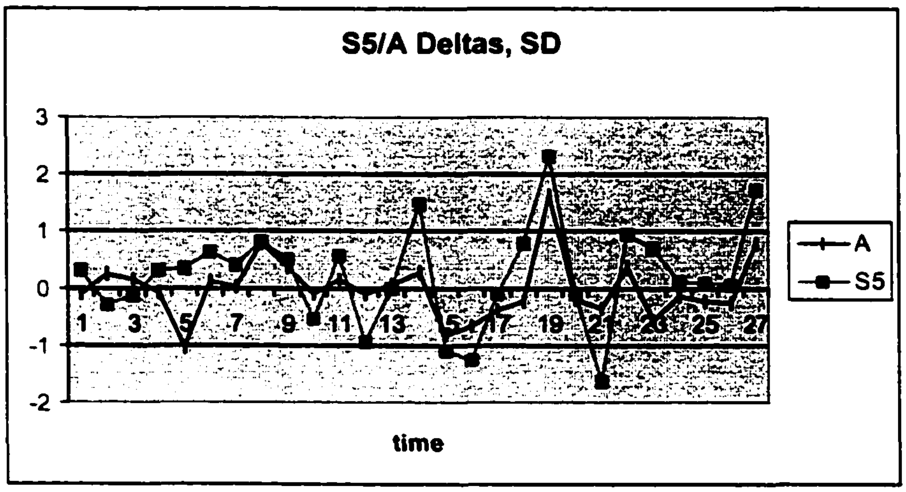


Figure 23. S5 with peers, deltas, SDs.

These correlations were due to the one significant individual interview S5 had for this measure, her interview with A, the female peer: $r = 0.680$, $p < 0.01$, $n = 27$.



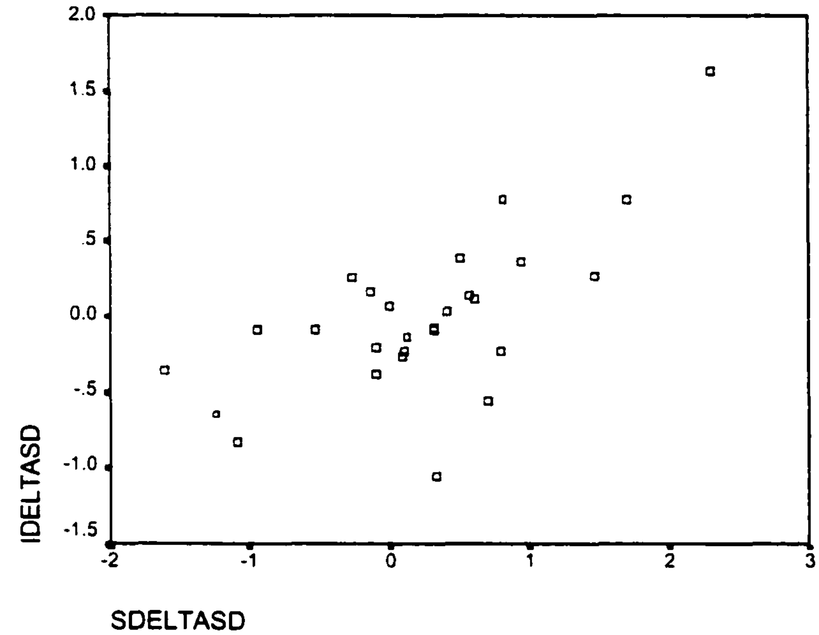
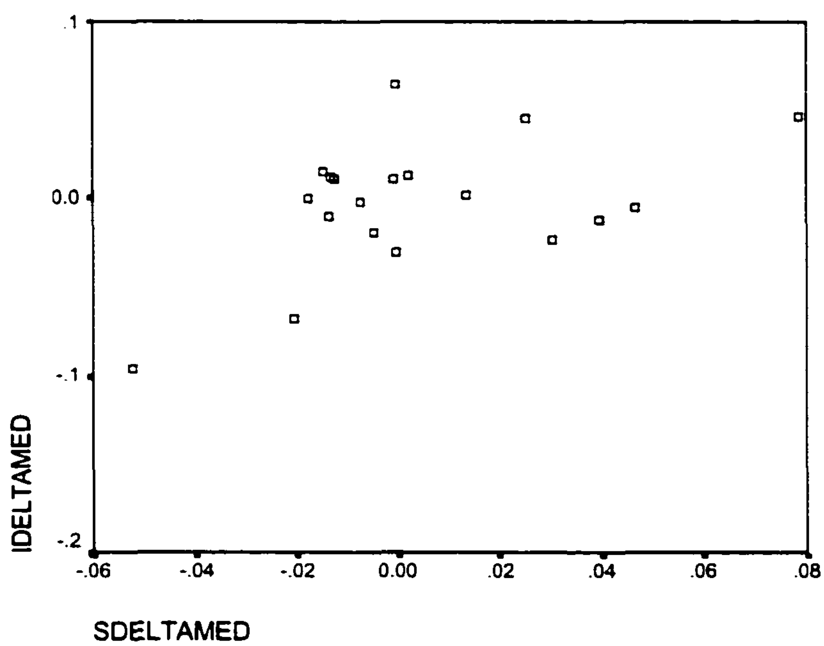


Figure 24. S5 with A, deltas, SDs.

Thus, for S2 and S6, there was no evidence of accommodation in the available data for the relevant measures, and for S7 and S5, there was one significant correlation in a relevant interview/measure each, and S5's finding was for the related delta measure, rather than the normalized score.

There were seven subjects who had significant results based on interlocutor gender in experiment one. While S1 and S10 had higher median pitches with females, S4 and S9 had larger SDs with females, and S4, S8, and S12 had larger ranges with females. S2 had higher median pitches with males. Neither S1 nor S10, the two subjects who had higher median pitches rather than more variable pitches with females, showed evidence of accommodation in their relevant interviews for their normalized scores for median pitch;

their normalized median pitches did not correlate with those of their male or female interlocutors, grouped or individually. However, both S1 and S10 each did have one significant positive correlation for the delta measure for median pitch, S1 with J, the male professor: $r = 0.495$, $p < 0.05$, $n = 19$, and S10 with P, the female professor: $r = 0.382$, $p < 0.05$, $n = 29$.



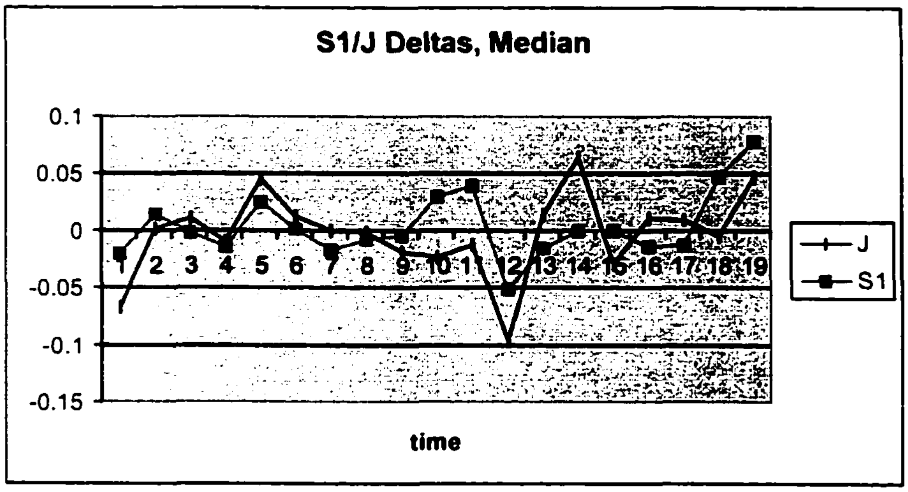
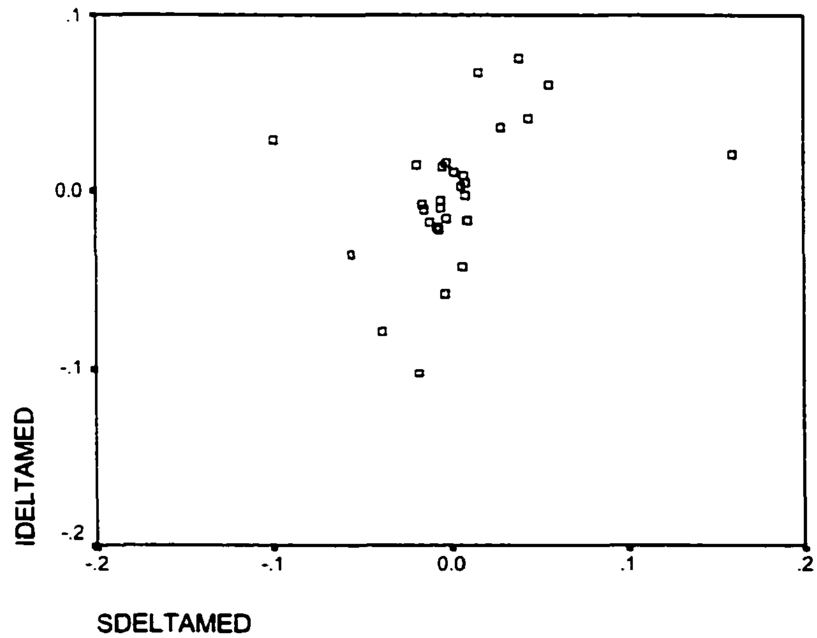


Figure 25. S1 with J, deltas, medians.



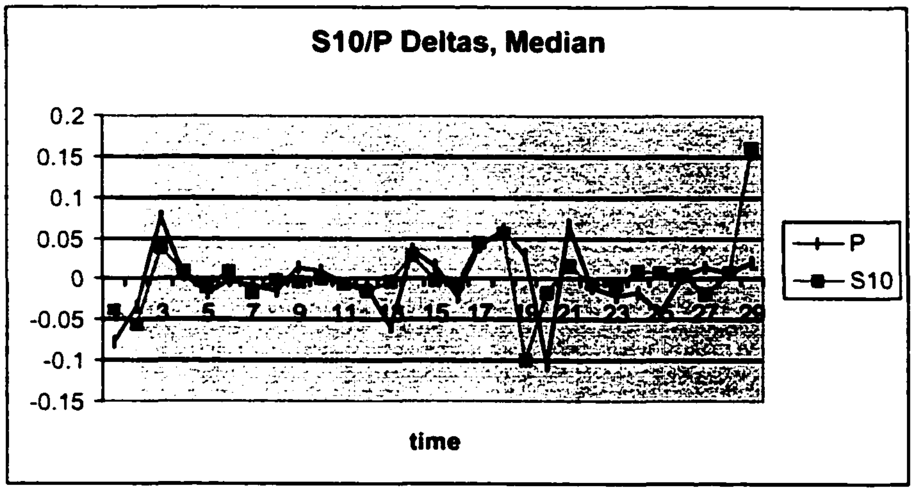


Figure 26. S10 with P, deltas, medians.

Furthermore, S2, the one subject who had higher pitches with males, and the only subject with significant results based on both interlocutor gender *and* status, did have a significant correlation for normalized median pitch with a male interlocutor, J, the male professor: $r = 0.698$, $p < 0.01$, $n = 21$. She also had a positive correlation with the professors as a group, but this finding reflects the correlation with the male professor. (Recall that the interviews between S2 and the peers were unanalyzable, due to poor sound quality for the interlocutors.)

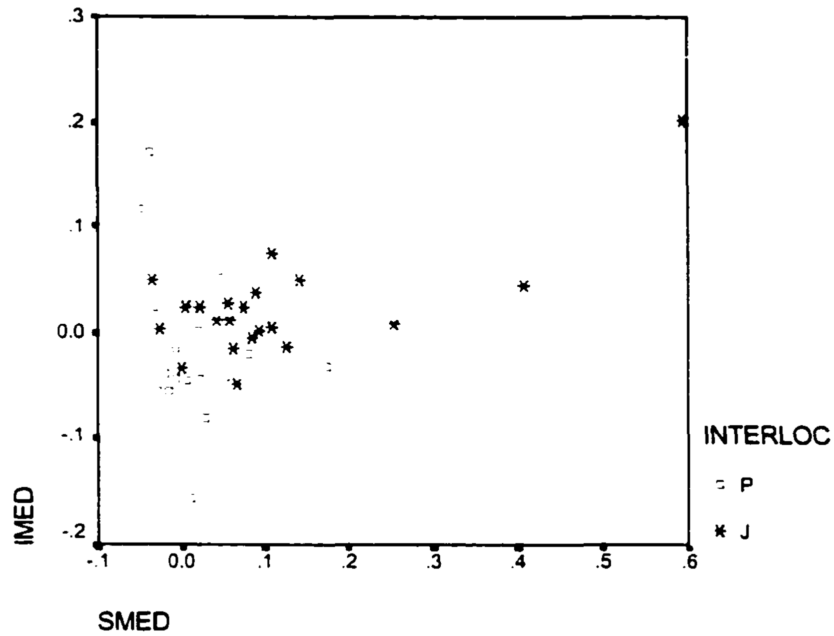
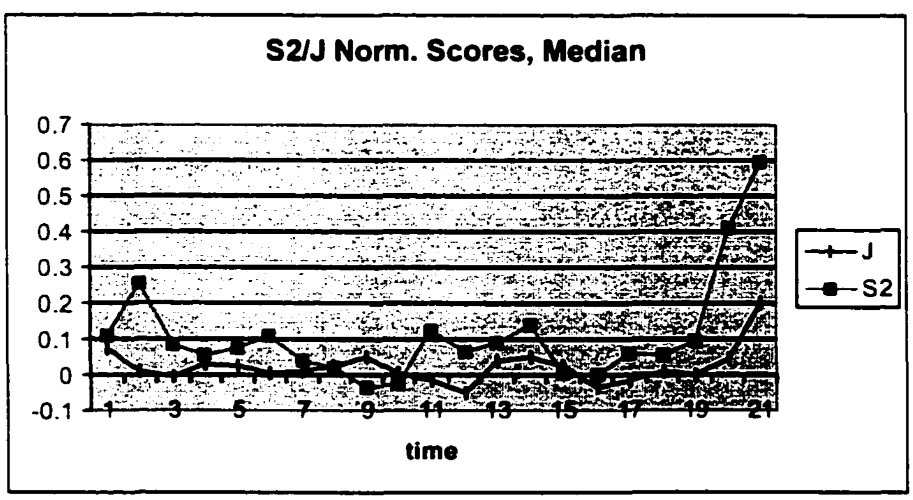


Figure 27. S2 with professor interlocutors, normalized scores, medians.



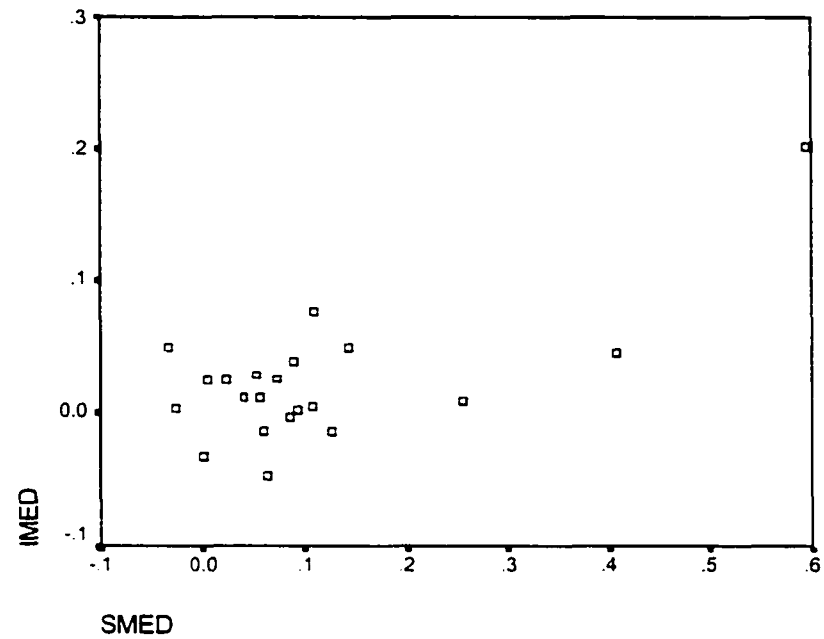


Figure 28. S2 with J, normalized scores, medians.

In summary, neither of the two subjects with higher median pitch with females had a significant correlation for normalized median pitch in the relevant measures/interviews, but they both did have a correlation for the delta measure. Furthermore, the one subject with higher median pitches with males had a significant correlation for normalized median pitch. The three subjects with gender-based results in experiment one who did not match the others who were more variable with women all had some kind of significant correlation in experiment two.

Some but not all of the subjects who had higher variability with females (none had higher variability with males) had some significant correlations in their interviews, as well. S8 (larger range) and S9 (larger SD) had no relevant significant correlations, but S4 and S12 did. S4 had both larger ranges and larger SDs with females in experiment one. In experiment two, her pitch range data positively correlated with her interlocutors when all four interviews were grouped together, and when the two male interlocutors were grouped together. However, again, these correlations may well be due to phenomena other than accommodation. The correlation between S4's ranges and the males' (B and J) ranges seems to be due to B's use of larger ranges than J's. B's points above J's, with or without P's (the female professor) rounding them out, create a linearly clustered cloud of points. (S4's interview with A, the female peer, was unavailable.)

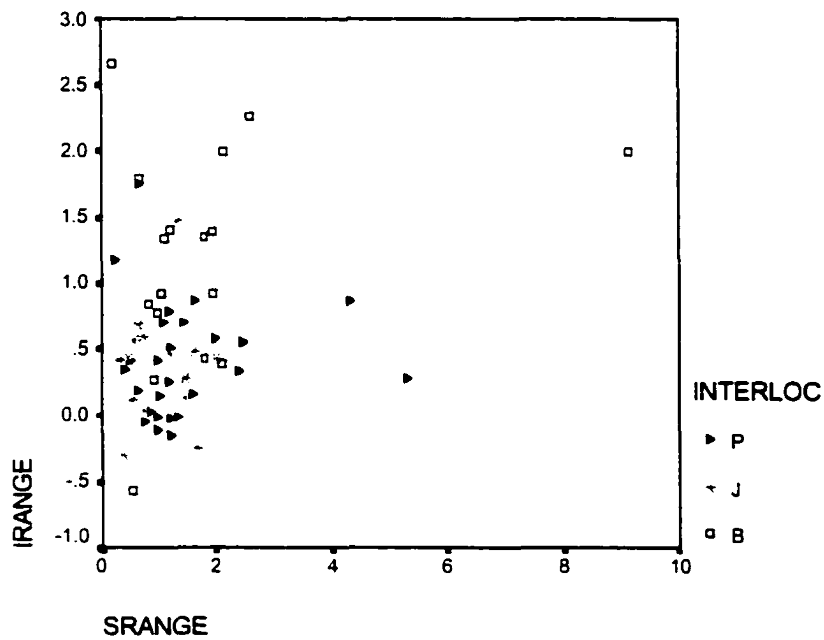


Figure 29. S4 with all three interlocutors, normalized scores, ranges.

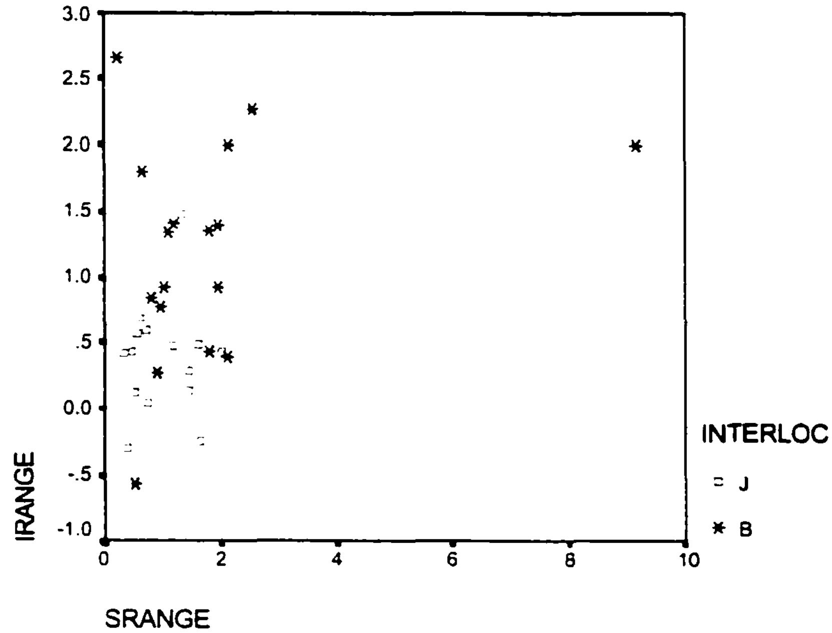
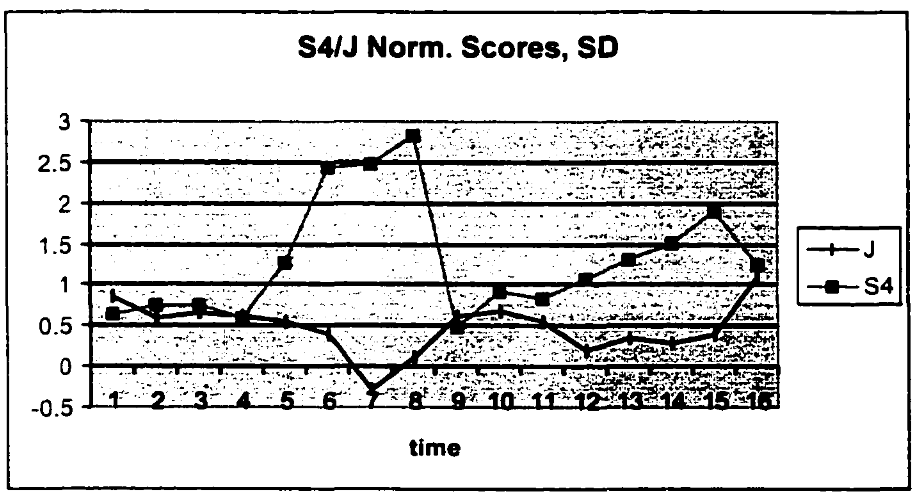


Figure 30. S4 with males, normalized scores, ranges.

It does appear that S4's data points with B are slightly higher than her points with J, suggesting slight global accommodation on S4's part. As with S7's median pitch data, these data reflect the situation where it may be impossible to distinguish between a kind of global (not involving co-variation) accommodation in pitch range and a kind of expectation-based social adjustment based on the perception of the interlocutor. It may be that S4 and the interlocutors in these conversations were adjusting their pitch ranges according to who (they thought) their interlocutor was and what they thought was appropriate in some way, or it may be that there was slight global accommodation on the parts of the interlocutors and S4. Evidence for the latter interpretation comes from the fact that S4 had her lowest normalized score (weighted average) for the entire interview for pitch range with the male professor and her highest range with the male peer (with the range with the female professor in between), and the male professor had his third smallest range measurement (out of 12 interviews) with S4, while the male peer had his second largest range (out of 10 interviews) with her. (See Appendix 3 for normalized scores calculated for the interviews as wholes for all subjects and interlocutors.) It may be that this small level of accommodation did not reach statistical significance in the individual interviews due to the small number of data points and small amounts of variation within the individual speakers' data and the relative weakness of the correlation; none of S4's individual interviews involved significant correlations for pitch range. However, even if the accommodation interpretation is correct, S4 having a larger pitch range with B versus J would not account for her experiment one result of having larger ranges with females than males.

S4 also had larger standard deviations of pitch with females in experiment one. However, the only significant correlation for SD she had in experiment two was a negative correlation with J, the male professor.



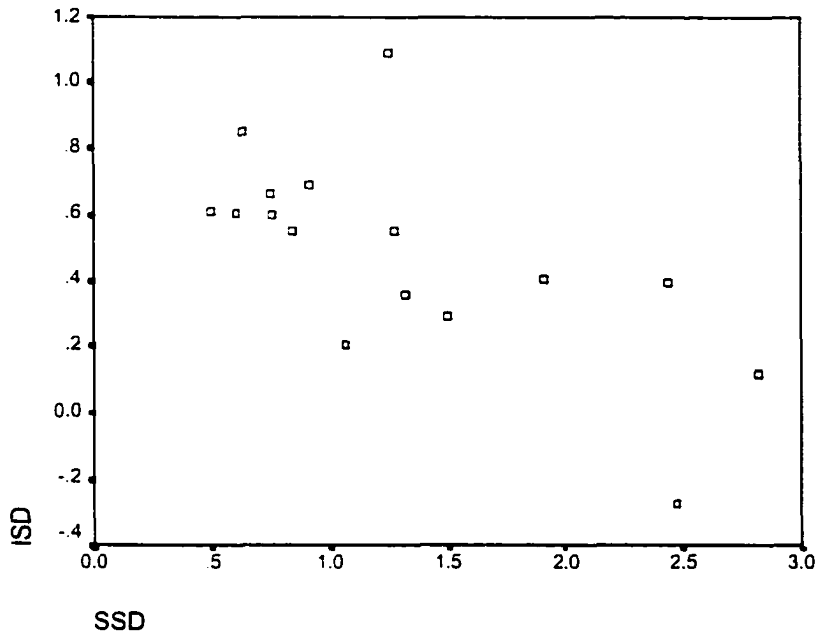
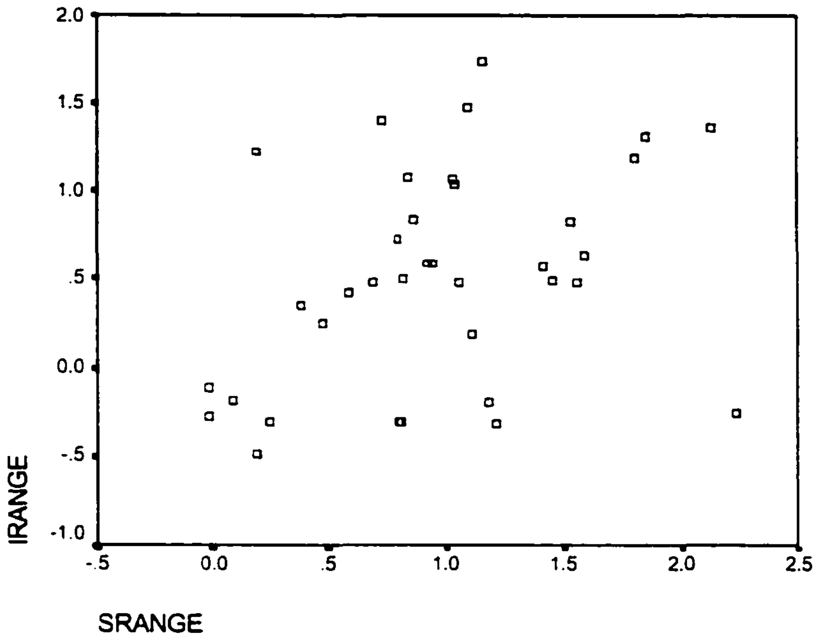


Figure 31. S4 with J, normalized scores, SDs.

This negative correlation probably does not explain S4's larger SDs with females in experiment one, although it may have contributed slightly to small SDs with males, as J had more larger SDs than smaller in this interview.

The only other subject with higher pitch variability with females in experiment one who had a relevant result in experiment two is S12, who had larger ranges with females. She had one significant correlation for normalized scores for range, in her individual interview with A, the female peer.



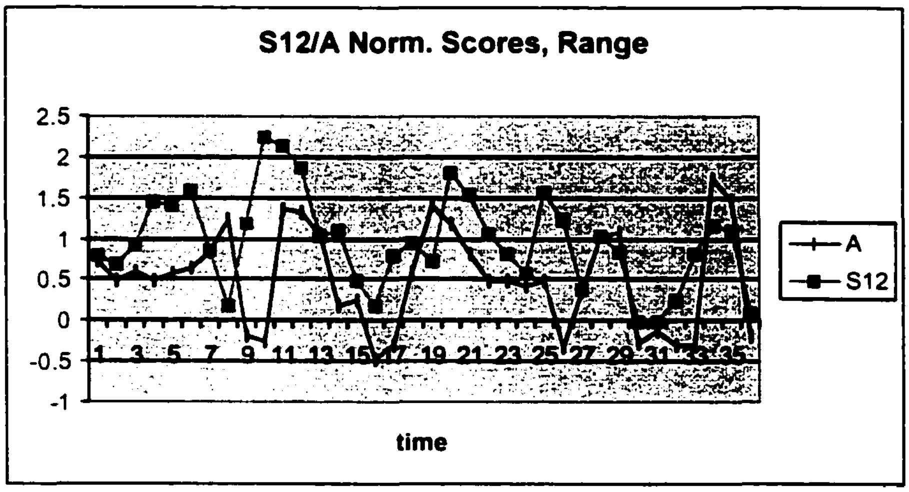


Figure 32. S12 with A, normalized scores, ranges.

She did not, however, have a significant correlation with P, the other female (professor) interlocutor.

Overall, the evidence for accommodation being responsible for the results of experiment one is not very strong. The one suggestive pattern is that the three "marked" gender-based subjects in experiment one (S1, S10, and S2) all showed some evidence of accommodation. It may be that accommodation was more responsible for S1, S10, and S2's results than it was for the others, but this interpretation is tentative, as some of the other gender-based subjects did show evidence of accommodation and some didn't. Considering all subjects with results in experiment one, for a few particular instances, it may have played a role, but many measures involved in experiment one results showed no signs of accommodation as measured. It seems that individuals use one or more strategies from an array of options. As the differential use of politeness strategies found amongst the subjects in experiment one showed, different individuals use different strategies, and one individual may use different strategies on different occasions or in different situations (i.e., with different interlocutors) or even combinations of them to accomplish their goals. However, there is much more to the accommodation story. The next section presents a summary of all of the accommodation-related results and the patterns within them.

3.2.2. Overall Accommodation Results

In order to obtain a comprehensive picture of how the 12 subjects may have accommodated to their interlocutors and vice versa, Pearson correlations were measured in SPSS for each measure, i.e., for normalized scores and deltas for median, SD, and pitch range, for each of the 43 interviews where both channels had measurable data. The level of correlation was also measured for several different interviews grouped together for each subject: with all four interlocutors as a group, and with female, male, peer, and professor interlocutors. For the interlocutors, interviews with all of the subjects were grouped together. When broken down into different measurement categories (median vs. SD vs. range, normalized score vs. delta, and positive correlation vs. negative correlation) and when grouped into categories based on gender and status of the interlocutor, several patterns emerge, so the data will be presented in several ways. It is important to keep in mind that significant correlations are being hypothesized to reflect minute-by-minute or real-time accommodation, as correlation indicates co-variation. They will not generally uncover what has been termed global accommodation that does not involve co-variation. As mentioned above, when more than one interview is grouped together, global accommodation may be indicated, but it is difficult to separate from expectation-based pitch adjustments.

It is again important to acknowledge that running a large number of correlations on the same data is problematic, in that a certain number of seemingly significant findings might occur just due to chance. Furthermore, even when the number of significant results is larger than chance could explain, it will not be clear which ones are not due to chance. The significant findings are reported below with this caveat in mind.

3.2.2.1. Group Correlations

As mentioned above, significant correlations for pooled interviews are often the result of one or more highly correlated individual interviews within the group. Other possibilities are either global accommodation or expectation-based pitch adjustments. As a result, these results looked at as a whole are less revealing, but they will be presented for completeness' sake. I will list them by subject and present interpretations for each possible grouping.

For interviews with all four interlocutors pooled together, there were ten significant correlations out of a possible 72 (12 subjects and six possible measures, i.e., normalized scores and deltas for median pitch, SD of pitch, and pitch range). Seven subjects had significant correlations for normalized scores for median pitch: S2, S4, S5, S7, S9, S11, and S12. There were only three significant results not based on this measure, S4's normalized scores for range and S5's deltas for SD and range.

	PEARSON CORRELATION COEFFICIENT (r)	SIGNIFICANCE LEVEL
S2 (median[12])	0.373, n = 41	0.016
S4 (median)	0.333, n = 59	0.010
S4 (range)	0.258, n = 59	0.048
S5 (median)	0.240, n = 104	0.014
S5 (delta SD)	0.233, n = 99	0.020
S5 (delta range)	0.227, n = 99	0.024
S7 (median)	0.209, n = 100	0.037
S9 (median)	0.250, n = 104	0.011
S11 (median)	0.301, n = 95	0.003
S12 (median)	0.221, n = 103	0.025

Table 6. Significant correlations for individuals with all four interlocutors.

The correlations for median pitch for S2 and S7, for range for S4, and for delta SD for S5 were described above in regard to their relevance to experiment one results (figures are presented above, as well); they appear to be due to the effects of individual correlations within the groups, and for S4 and S7, possibly to either global (not involving co-variation) accommodation or expectation-based results. It is very difficult to tease apart these two interpretations, as the distinction is in the mind of the speaker, rather than

[12] Normalized scores are being reported, unless a measure is labeled "delta."

in her behavior, i.e., whether her pitches are a reaction to the pitches she is perceiving or a reaction to other social stimuli in the situation.

Several of the other results have similar explanations. For instance, many of them are due to individual interview correlations within the groups: S12 had individual median correlations with B and P, S9 does with A and B, and S5 has an individual correlation for delta range with A. (Recall that individual interlocutors, referred to by their first initial, are as follows: A = female peer, B = male peer, P = female professor, and J = male professor.)

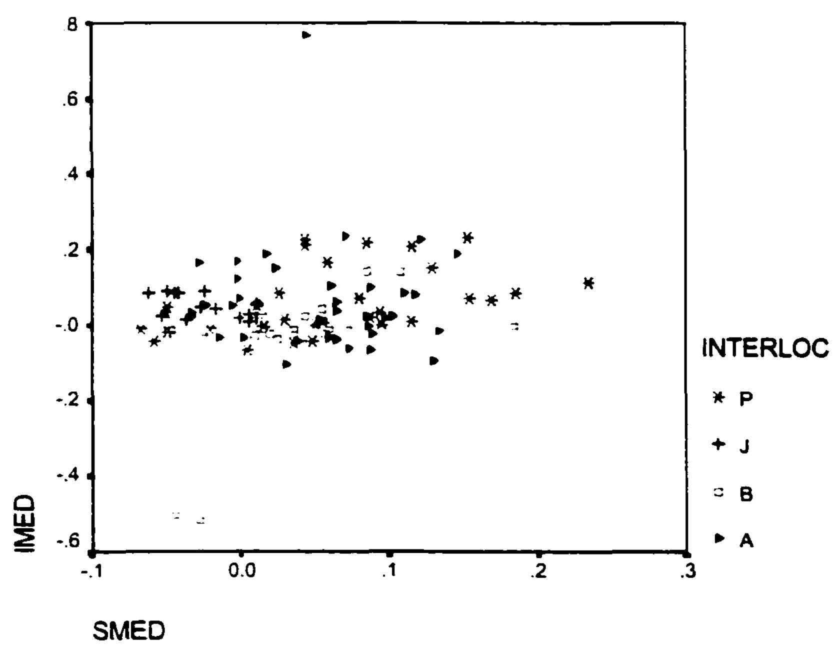


Figure 33. S12 with all four interlocutors, normalized scores, medians.

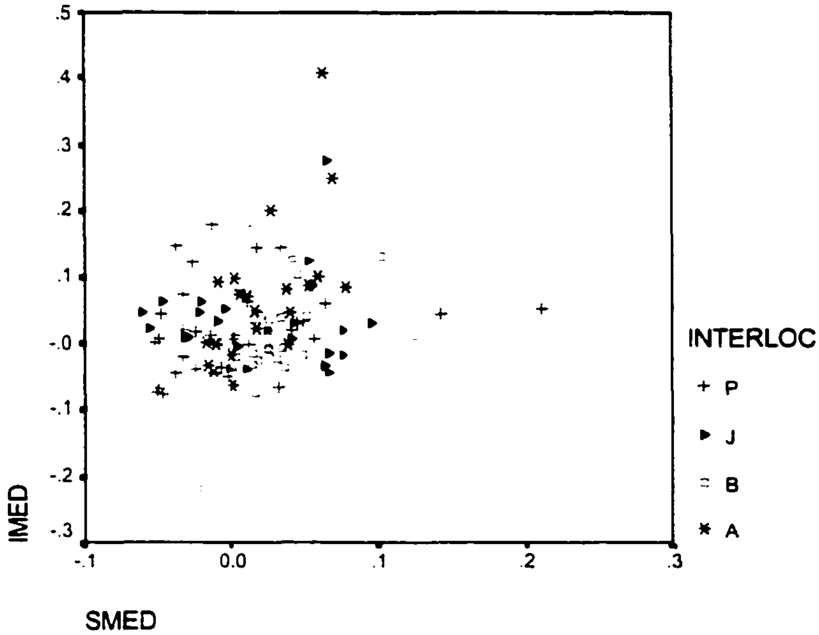


Figure 34. S9 with all four interlocutors, normalized scores, medians.

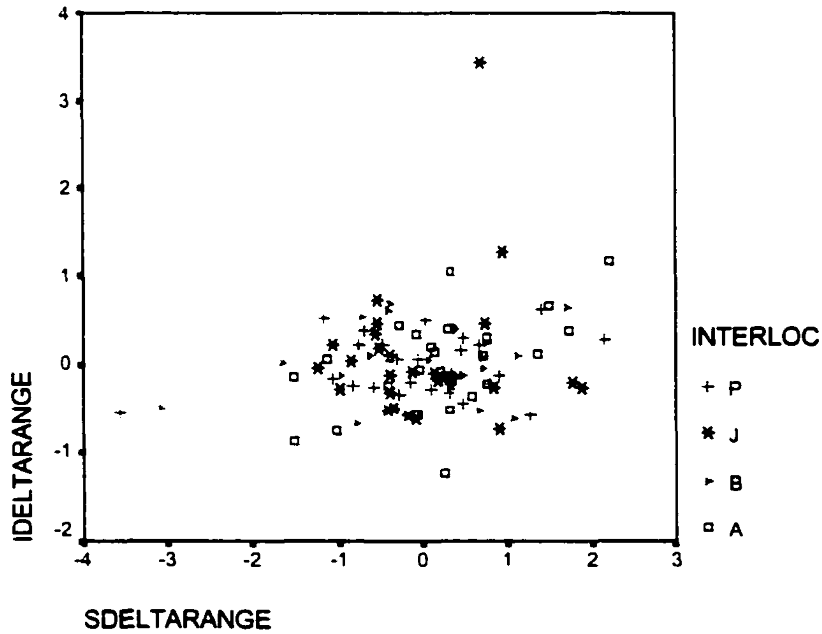


Figure 35. S5 with all four interlocutors, deltas, ranges.

S4 also has an individual correlation with B that may explain her group median significant result, but it may additionally be the case that a non-significant correlation with J's median pitches are contributing as well.

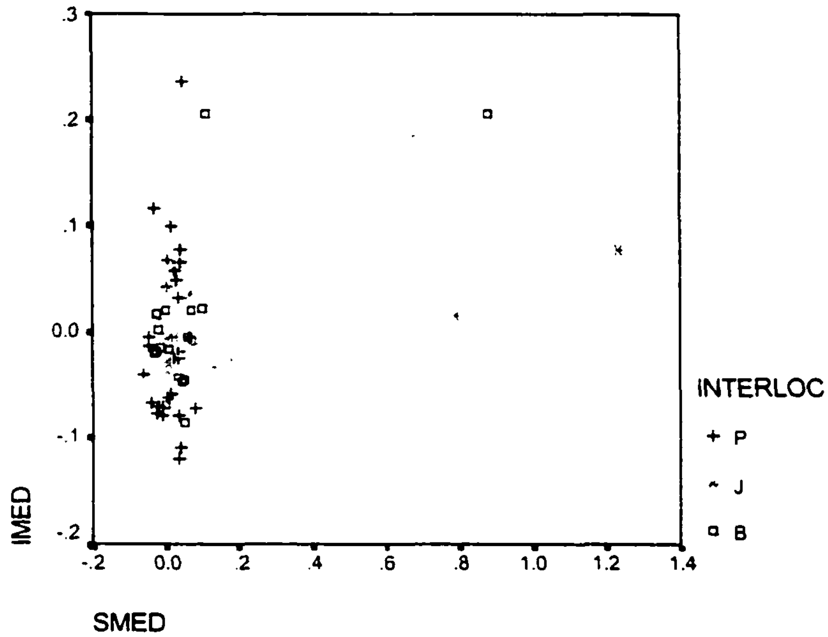


Figure 36. S4 with all three interlocutors, normalized scores, medians.

S5's group correlation for median pitch may be at least in part due to her individual correlation with P. It may also be due to J's pitches being in general higher with her than P's (for their normalized scores). This pattern does not appear to be accommodation, however, as S5's median pitches are not in general higher with J than they are with P according to both the data presented in the following scatterplot, and according to S5's overall scores for each interview, presented in Appendix 3.

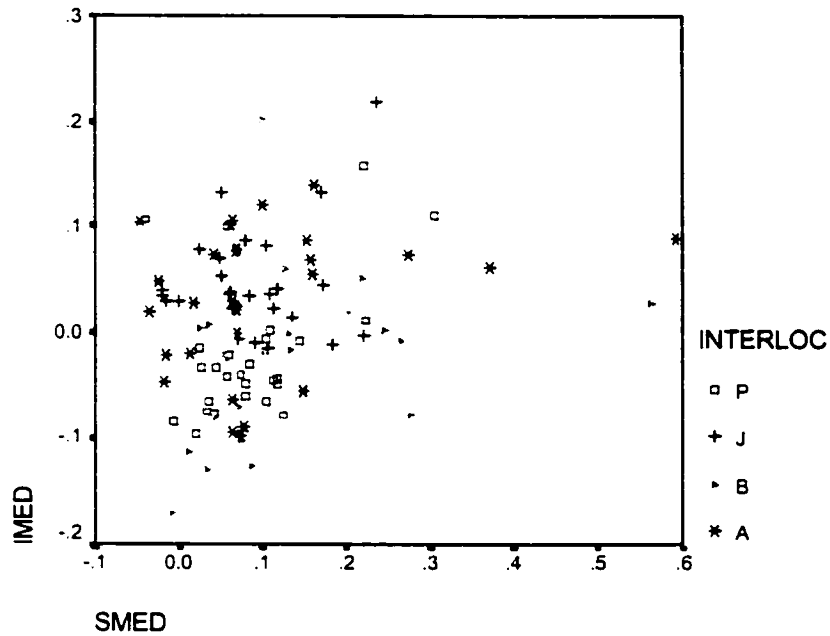


Figure 37. S5 with all four interlocutors, normalized scores, medians.

S11's median pitch data with all four interlocutors also appears to be due to a combination of an individual result (with B) and a difference between interlocutors' pitches with her. P's pitches are in general higher than B's. However, unlike for S5, there is some evidence for global accommodation, as S5's pitches are in general higher with P than they are with B. The overall interview scores are consistent with this interpretation, as well. S11 has a higher overall normalized median pitch score for her interview with P than with B, and P has her highest normalized score with S11 (out of 12 interviews) while B has his second lowest score (out of 10, see Appendix 3).

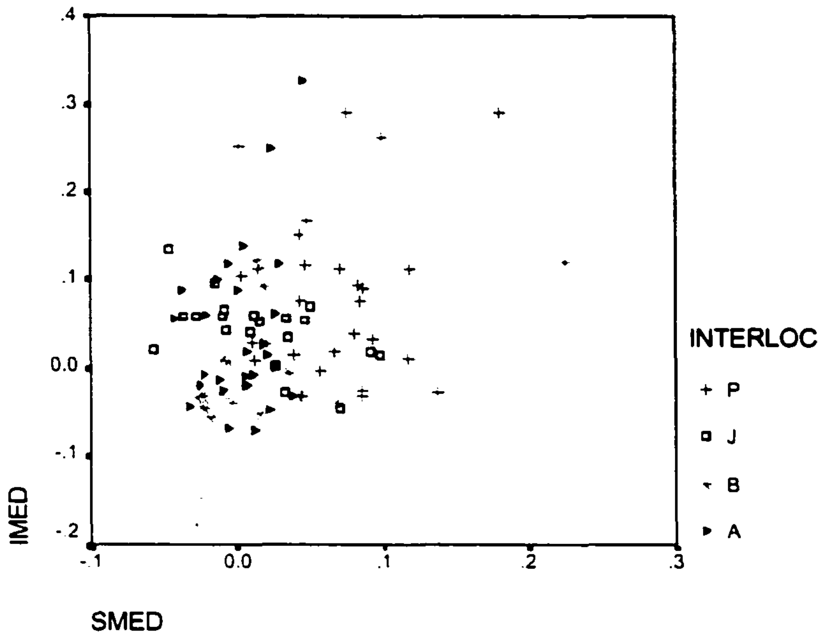


Figure 38. S11 with all four interlocutors, normalized scores, medians.

Keeping these interpretations in mind, there is a pattern amongst the groups of significant correlations for subjects with all four interlocutors. They are very largely (seven out of ten) for normalized scores for median pitch.

Correlations were also measured for each interlocutor with all subjects grouped together. With all nine subjects, A had significant positive correlations for median ($r = 0.190$, $p < 0.01$, $n = 222$), delta SD ($r = 0.190$, $p < 0.01$, $n = 213$), and delta range ($r = 0.177$, $p = 0.01$, $n = 213$). With all ten subjects, B had significant positive correlations for median ($r = 0.189$, $p < 0.01$, $n = 208$) and SD ($r = 0.155$, $p < 0.05$, $n = 208$). With all 12 subjects, P had significant positive correlations for median ($r = 0.173$, $p < 0.01$, $n = 334$)

and delta median ($r = 0.133$, $p < 0.05$, $n = 322$). These correlations are doubtlessly due to individual interview correlations within the groups. Interestingly, J had no significant correlations for all 12 subjects grouped together.

There were also several significant results based on gender groupings and status groupings. Out of a possible 114 (54 possible with females as a group and 60 possible with males as a group, i.e., multiplying the numbers of groups of interviews with each subject by the six measures), there were 15 significant correlations based on interlocutor gender. For interlocutor status, out of a possible 126 (54 possible with peers and 72 possible with professors as a group) there were 17. The significant results are listed by subject below.

	MEASURE/INTERLOC. GROUP	PEARSON COEFFICIENT	SIGNIFIC. LEVEL
S2	median, profs.[13]	r = 0.373, n = 41	0.016
S3	SD, peers	r = - 0.400, n = 28	0.035
	range, females	r = - 0.332, n = 45	0.026
	delta median, females	r = 0.338, n = 43	0.027
	median, males	r = - 0.433, n = 27	0.024
S4	median, males	r = 0.485, n = 33	0.004
	range, males	r = 0.354, n = 33	0.043
S5	median, profs.	r = 0.294, n = 54	0.031
	delta SD, peers	r = 0.441, n = 47	0.002
	delta range, peers	r = 0.364, n = 47	0.012
	median, females	r = 0.314, n = 55	0.020
	delta SD, females	r = 0.416, n = 53	0.002
	delta range, females	r = 0.373, n = 53	0.006
	range, males	r = 0.318, n = 49	0.026
S6	delta range, peers	r = 0.337, n = 36	0.044
S9	median, peers	r = 0.444, n = 47	0.002
	median, females	r = 0.334, n= 54	0.013
	range, females	r = 0.318, n = 54	0.019
	delta median, females	r = 0.329, n = 52	0.017
S10	delta SD, profs.	r = 0.355, n = 45	0.017
	delta range, profs.	r = 0.350, n = 45	0.019
	SD, males	r = 0.399, n = 36	0.016
S11	SD, profs.	r = 0.499, n = 47	0.000
	range, profs.	r = 0.340, n = 47	0.020
	delta SD, profs.	r = 0.321, n = 44	0.034

[13] This result was also reported above for results with all interlocutors, as the interviews with professors were the only two interviews available for this subject.

	median, peers	r = 0.318, n = 48	0.028
	median, females	r = 0.278, n = 55	0.040
	SD, males	r = 0.394, n = 40	0.012
S12	median, profs.	r = 0.379, n = 45	0.010
	SD, profs.	r = - 0.440, n = 45	0.002
	range, profs.	r = - 0.316, n = 45	0.034
	delta median, profs.	r = 0.375, n = 43	0.013

Table 7. Significant results for individuals with different interlocutor groups.

These group-based results have explanations that are very similar to those presented above for the results for interviews with all four interlocutors pooled together. They are largely explained by individual correlations within the groupings. All of the relevant scatterplots are given in Appendix 4.

There are some patterns in the above group-based results. Out of the 15 gender-based group results, nine were with the two females as a group, and six were with the two males as a group. Out of the 17 status-based group results, six were with the two peers as a group versus 11 with the professors. More group results occurred with females and professors than with males and peers. There are 21 normalized score group results and 11 delta group results, and 13 median, nine SD, and 10 range group results (combining normalized scores and deltas). The most common kind of result was normalized median score, with ten significant results. For all of the different kinds of (groups of) interlocutors, the significant results were spread relatively equally over the different measurements, except for the male interlocutors, where there were no significant results based on any of the delta measures. These patterns are present, but it is important to recall that the significant correlations for these pooled groups are largely based on results from individual interviews. The patterns present in the pooled groups are most likely explained by the individual correlations discussed in the next section. It is any patterns found there that must be explained.

3.2.2.2. Individual Correlations

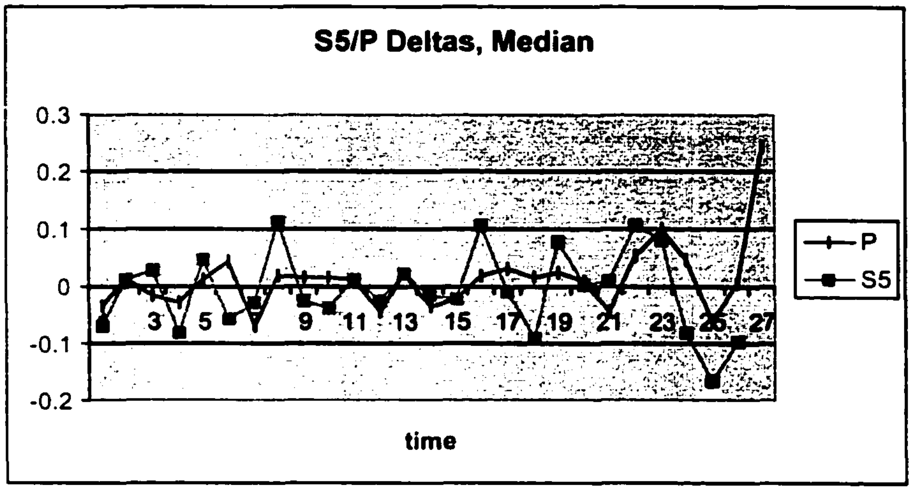
Out of 258 possible places for correlation (six measures times 43 individual interviews), 39 significant correlations were found. The overall breakdown amongst these individual correlations looked at together is: Thirty-one of these were positive and eight were negative: 21 were for median pitch, eight were for standard deviation of pitch, and ten were for 80% pitch range; 21 were for normalized scores and 18 were for deltas. The specific measurement breakdown is normalized score and delta, median pitch, 12 and nine, respectively, normalized score and delta, SD, five and three, and normalized score and delta, pitch range, four and six. The kind of measure with the most significant results was thus median pitch. Another rather striking pattern from all of these significant individual results pooled together is that six out of eight negative scores are for median pitch. All of significant correlations from individual interviews are listed below by subject.

	MEASURE/INTERLOCUTOR	PEARSON COEFFICIENT	SIGNIFIC. LEVEL
S1	delta median, J	r = 0.495, n = 19	0.031
	delta range, P	r = 0.590, n = 14	0.026
S2	median, J	r = 0.698, n = 21	0.000
S3	SD, B	r = - 0.632, n = 13	0.021
	delta median, P	r = 0.432, n = 29	0.019
S4	median, B	r = 0.686, n = 17	0.002
	SD, J	r = - 0.671, n = 16	0.004
S5	median, P	r = 0.440, n = 27	0.022
	delta median, P	r = 0.453, n = 26	0.020
	delta SD, A	r = 0.680, n = 27	0.000
	delta range, A	r = 0.521, n = 27	0.005
S6	delta range, B	r = 0.476, n = 18	0.046
S7	median, B	r = 0.416, n = 24	0.043
S8	median, J	r = - 0.542, n = 15	0.037
	delta median, J	r = - 0.623, n = 14	0.017
S9	median, A	r = 0.616, n = 21	0.003
	median, B	r = 0.419, n = 26	0.033
	range, A	r = 0.618, n = 21	0.003
	delta median, P	r = 0.398, n = 32	0.024
	delta range, A	r = 0.479, n = 20	0.033
S10	SD, J	r = 0.539, n = 17	0.025
	range, P	r = 0.411, n = 30	0.024

	delta median, P	r = 0.382, n = 29	0.041
	delta SD, P	r = 0.473, n = 29	0.010
	delta range, P	r = 0.503, n = 29	0.005
S11	median, B	r = 0.553, n = 20	0.012
	median, J	r = - 0.528, n = 20	0.017
	SD, J	r = 0.791, n = 20	0.000
	range, J	r = 0.551, n = 20	0.012
	delta median, J	r = - 0.585, n = 18	0.011
	delta SD, J	r = 0.737, n = 18	0.000
	delta range, J	r = 0.507, n = 18	0.032
S12	median, B	r = 0.535, n = 22	0.010
	median, J	r = - 0.561, n = 14	0.037
	median, P	r = 0.458, n = 31	0.009
	SD, A	r = 0.350, n = 36	0.036
	range, A	r = 0.371, n = 36	0.026
	delta median, A	r = - 0.429, n = 35	0.010
	delta median, P	r = 0.414, n = 30	0.023

Table 8. Significant correlations in individual interviews.

Pearson correlation coefficients measure linear correlation, whether it be positive or negative. The following scatterplots demonstrate the linear correlation when the measurements for a subject and interlocutor for an interview are plotted against each other. I will provide several representative examples from the significant correlations listed in the chart just above. The graphs for the significant individual interview results not shown here are all provided in Appendix 5. Scatterplots display the linear correlation present, but they do not preserve "real time." They do not present the data points as they occurred in time across the interview. To represent the data in real time, line graphs are also provided. Several positive correlations will be given, and then a few negative ones.



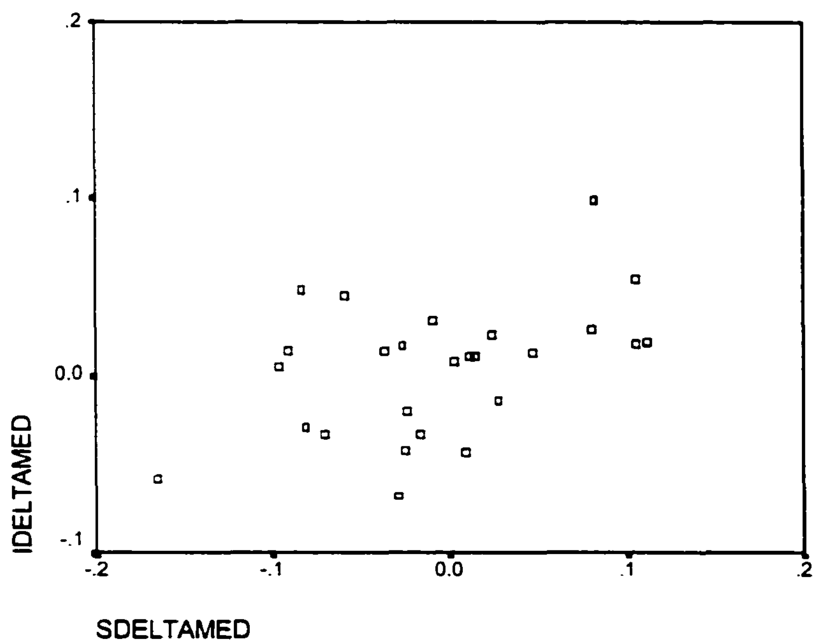
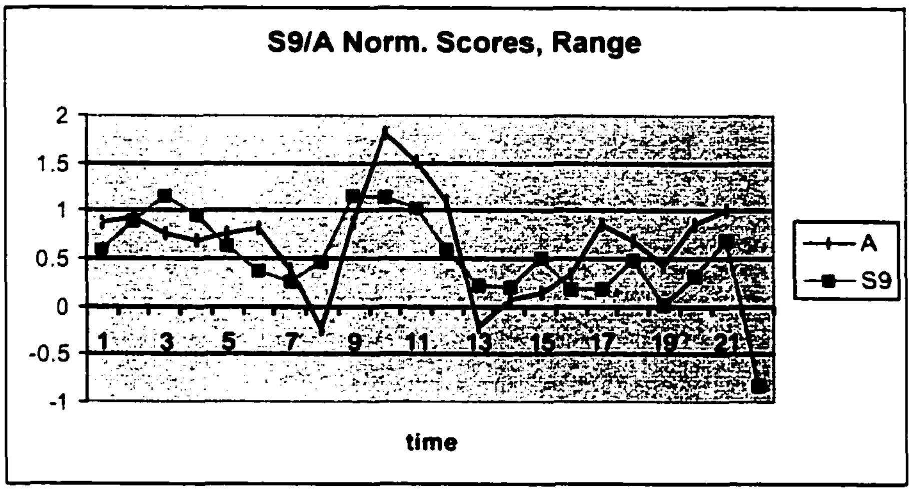


Figure 39. S5 with P, deltas, medians.



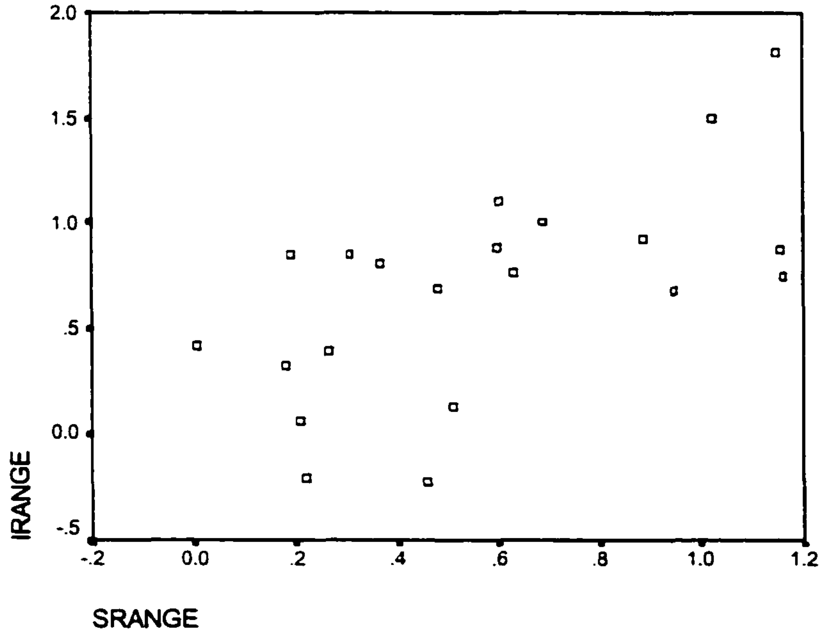
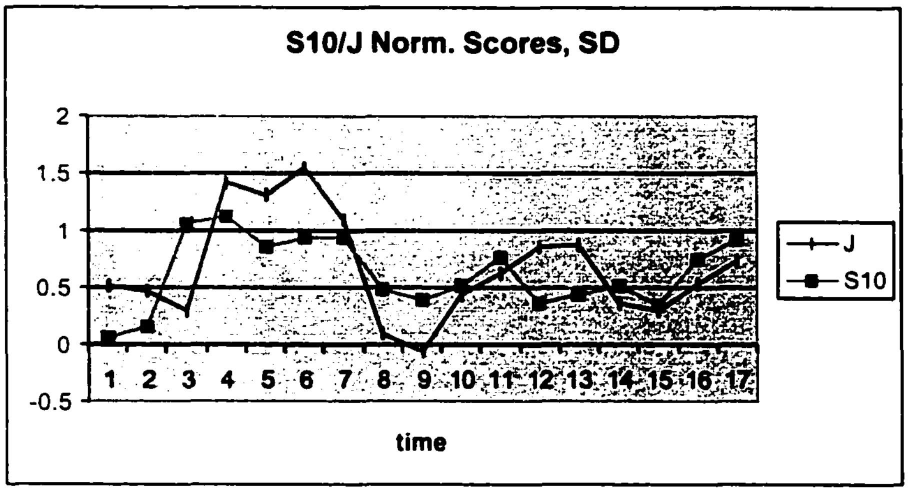


Figure 40. S9 with A, normalized scores, ranges.



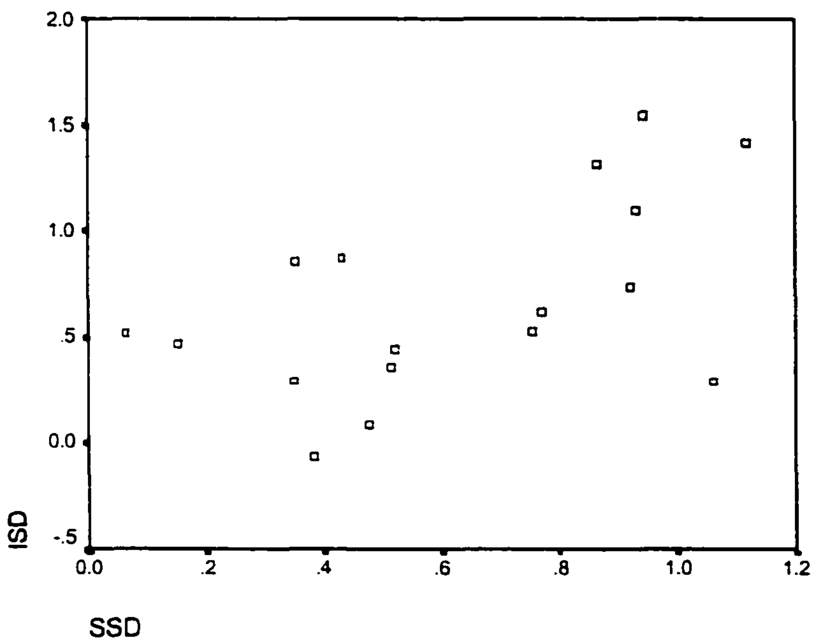
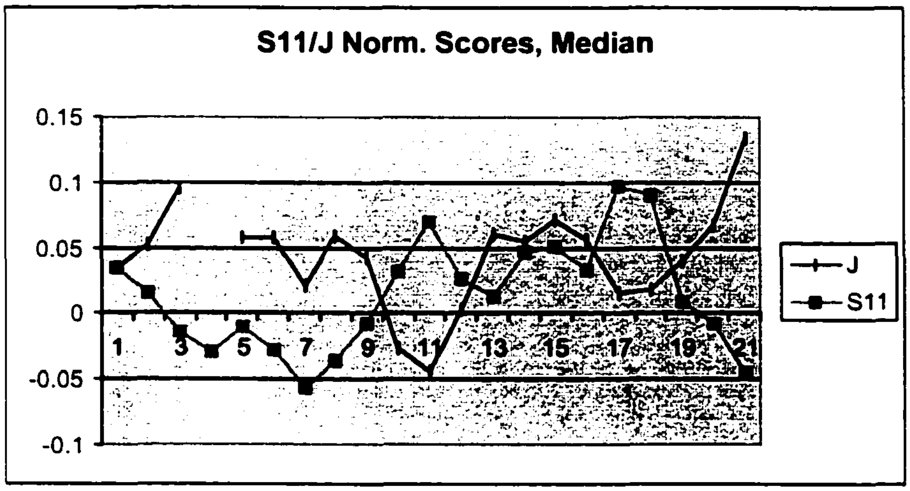


Figure 41. S10 with J, normalized scores, SDs.

Negative correlations:



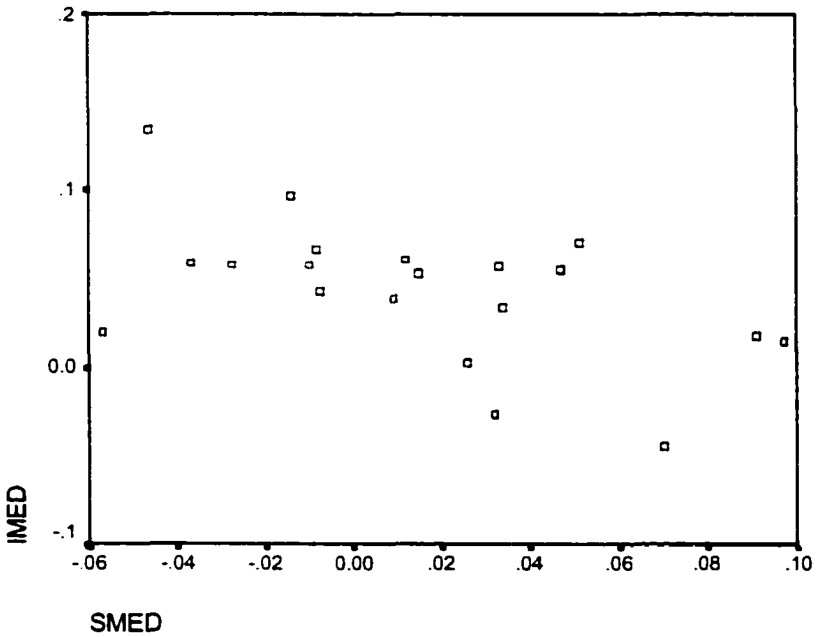
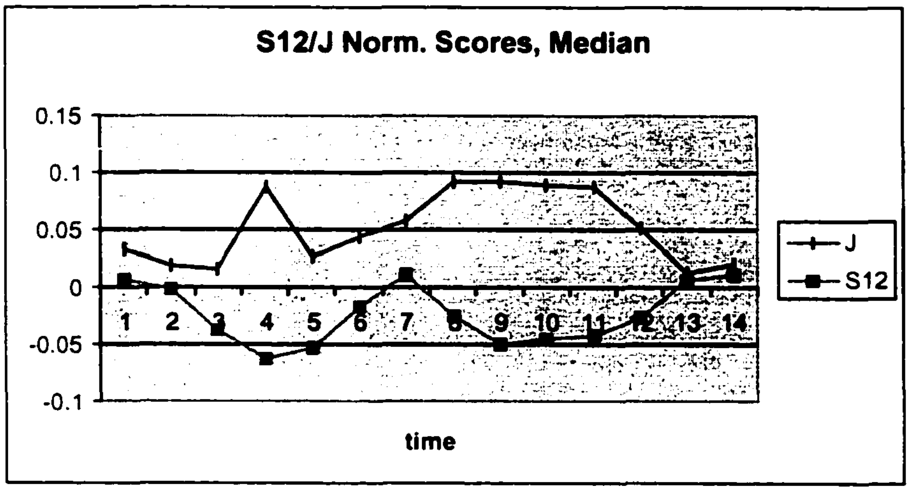


Figure 42. S11 with J, normalized scores, medians.



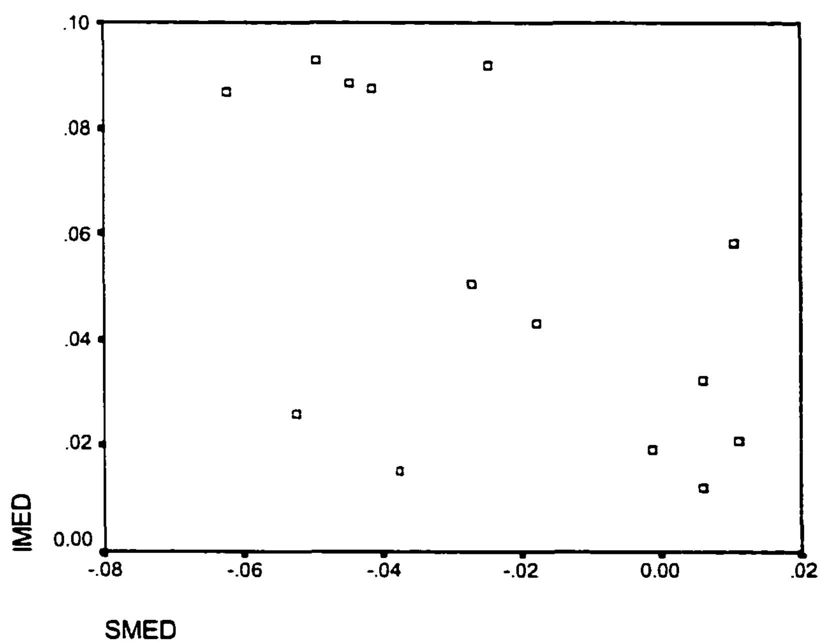


Figure 43. S12 with J, normalized scores, medians.

There are some patterns in the results broken down by subject. S5 only had results with females (four results) and S9 had mostly results with females (four with females, one with a male). However, two other subjects had more results suggesting accommodation with male interlocutors; S11 had all of her seven results with males, as did S4, although she only had two significant results. One subject, S10, had all her five significant results with professors. (S1 had both results reported here with professors, but recall that her interviews with peers were unavailable.) No subject had results only with peers. It appears that different subjects accommodate to different (types of) people. It is interesting that the patterns are based on the identity or characteristics of the interlocutor, rather than the kind of measure. As one of the stated goals of this study was to try to discover who is accommodated to, these interlocutor-based patterns will be explored further. The significant results by interlocutor are listed below.

		Normalized scores		Deltas	
Interlocutor		Positive	Negative	Positive	Negative
A (F peer):	Median	1	0	0	1
	SD	1	0	1	0
	Range	2	0	2	0
B (M peer):	Median	5	0	0	0
	SD	0	1	0	0
	Range	0	0	1	0
J (M prof.):	Median	1	3	1	2
	SD	2	1	1	0
	Range	1	0	1	0
P (F prof.):	Median	2	0	5	0
	SD	0	0	1	0
	Range	1	0	2	0

Table 9. Significant results for each interlocutor.

Several of the interlocutors have a prominent measure within their correlations. A had the most significant correlations for pitch range for both normalized scores and deltas, B had the most for normalized scores for median pitch, and P had the most for median pitch, especially delta median pitch. J's significant results were spread relatively equally across the different kinds of measures, but he had the most negative correlations.

In order to test whether the interlocutors' social characteristics of gender and status (perhaps in addition to personality or other factors that could be explaining the above individual interlocutor results) were playing a role in determining whether their interviews involved accommodation, the above individual results were grouped according to interlocutor gender and status.

		Normalized scores		Deltas	
Interlocutor Group		Positive	Negative	Positive	Negative
Females:	Median	3	0	5	1
	SD	1	0	2	0
	Range	3	0	4	0
Males:	Median	6	3	1	2
	SD	2	2	1	0
	Range	1	0	2	0
Peers:	Median	6	0	0	1
	SD	1	1	1	0
	Range	2	0	3	0
Professors:	Median	3	3	6	2
	SD	2	1	2	0
	Range	2	0	3	0

Table 10. Significant results for each interlocutor group.

These numbers can be totaled in different ways, revealing several patterns among the groups. Female interlocutors had 18 positive significant correlations compared to the males' 13, while males had seven negative significant correlations compared to the females' one, making totals of 19 for females and 20 for males. Peers had 13 positive and two negative correlations while professors had 18 positive and six negative, making totals of 15 and 24 for them. While males and females had very similar totals, there was a difference in the positive vs. negative correlation distribution; females had more positive results while males had more negative results, although the negative numbers were much smaller. For peers and professors, professors had higher numbers for both positive and negative correlations.

The groups had differences in the distribution among the three kinds of pitch measures, as well. Combining results based on normalized scores and deltas, females had nine results for median pitch (eight positive and one negative), three for SD (all positive), and seven for range (all positive), while males had 12 for median (seven positive and five negative), five for SD (three positive and two negative), and three for range (all positive). Males tended to have results based on median pitch, while females tended to have high numbers for both median and range. Peers had seven results for median pitch (six positive and one negative), three for SD (two positive and one negative), and five for range (all positive), while professors had 14 for median (nine positive and five negative), five for SD (four positive and one negative), and five for range (all positive). Peers tended to have results based on median and range, while professors tended to have them for median. In this dimension, females and peers pattern together as do males and professors. It is interesting to note that there were no negative correlations for range, and only two for SD; a measure of pitch tessitura is apparently much more likely to involve a negative correlation than a measure of pitch variability.

There were also differences based on normalized scores versus deltas. Females had seven significant results based on normalized scores (all positive) and 12 based on deltas (11 positive and one negative), while males had 14 based on normalized scores (nine positive and five negative) and six based on deltas (four positive and two negative). Thus, females were much more likely to have a significant result based on a delta measure, while males were much more likely to have one based on a normalized score measure.

Peers had ten significant results based on normalized scores (nine positive and one negative) and five based on deltas (four positive and one negative), while professors had 11 based on normalized scores (seven positive and four negative) and 13 based on deltas (11 positive and two negative). Peers were more likely to have a result based on a normalized score measure, and professors were *slightly* more likely to have one based on a delta measure.

With the previously mentioned caveat in mind regarding the possibility that some of the significant results reported and tallied just above are due to chance, multivariate analyses of variance were run in SPSS to test what differences in numbers of significant results between interlocutor groups, if any, were not due to chance. ANOVAs were also run to see if any of the variables involved interacted significantly and to test for any other effects on the numbers of significant correlations. It was indeed found that some of these differences between the different interlocutor groups were statistically significant, and in several cases, variables interacted. ANOVAs were run to test the effects of variables on numbers of significant correlations for positive and negative results considered separately (i.e., number of insignificant results versus number of positive results versus number of negative results) and together (i.e., number of significant results, whether they are positive or negative, versus number of insignificant results).

Firstly, the numbers of significant results were significantly affected by the kind of measure, median pitch, standard deviation of pitch, or pitch range. For positive and negative results considered separately, $F(2, 255) = 3.486$, $p < 0.05$, and for all positive and negative results considered together, $F(2, 255) = 4.545$, $p < 0.05$. Median pitch had the largest number of significant results.

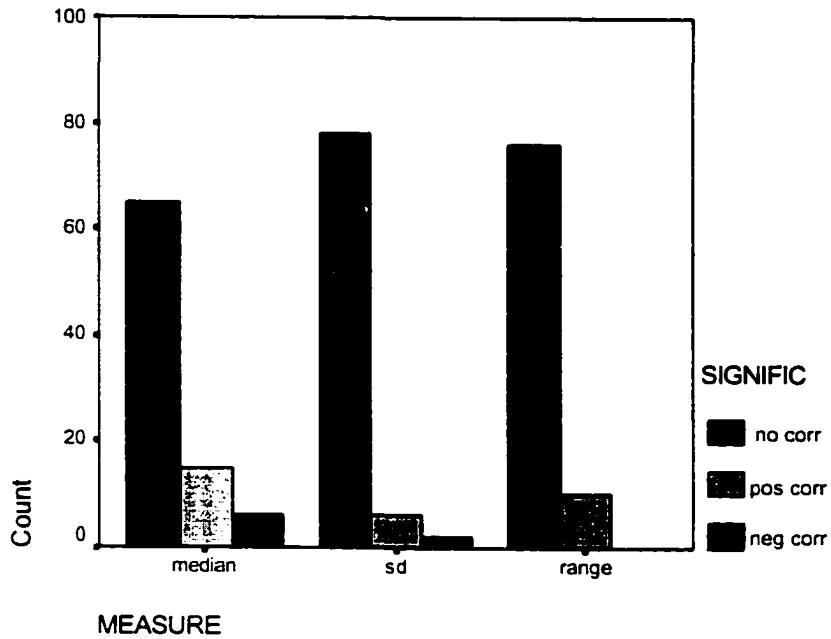


Figure 44. Measure by type of significant correlation (none, positive, negative).

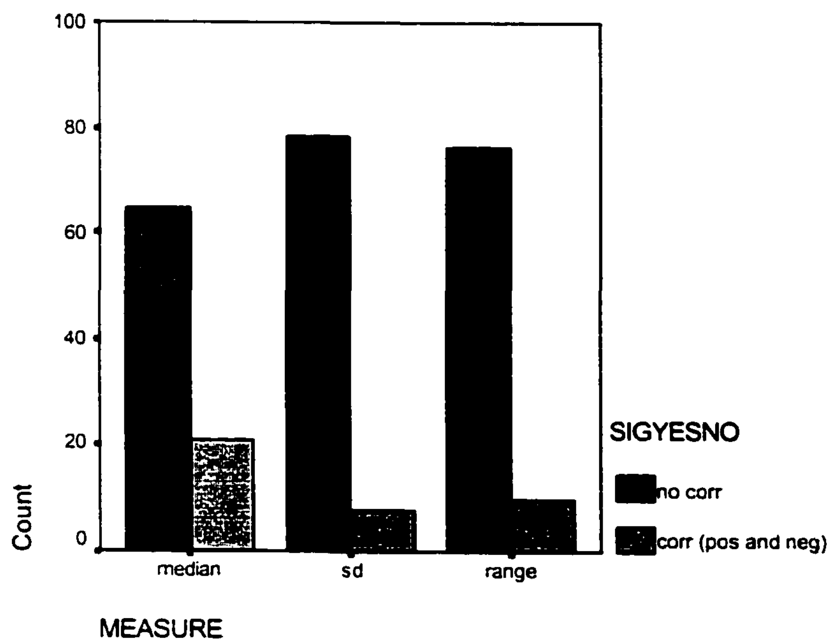
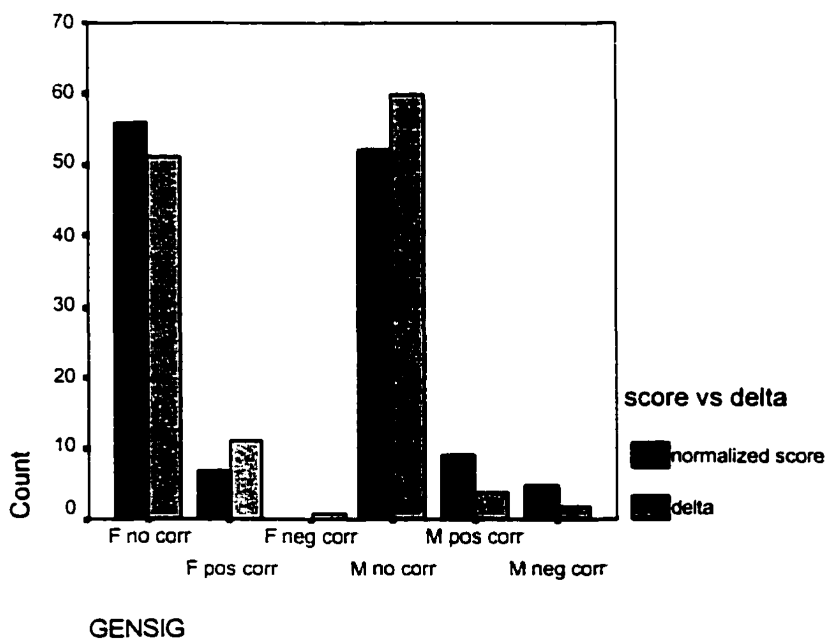


Figure 45. Measure by presence of significant correlation, positive and negative together.

The second kind of significant effect involved interlocutor gender interacting with type of measurement. This interaction happened in two ways. Interlocutor gender interacted with normalized score measure versus delta measure to significantly affect the number of (significant) results. For positive and negative results considered separately, $F(1) = 4.355$, $p < 0.05$, and for all positive and negative results considered together, $F(1) = 5.258$, $p < 0.05$. These results demonstrate that the differences described above that female interlocutors had more correlations based on delta measures and that male interlocutors had more based on normalized score measures were statistically significant.

Figure 46. Number of significant correlations for score vs. delta, by interlocutor gender and kind of



correlation (positive vs. negative).

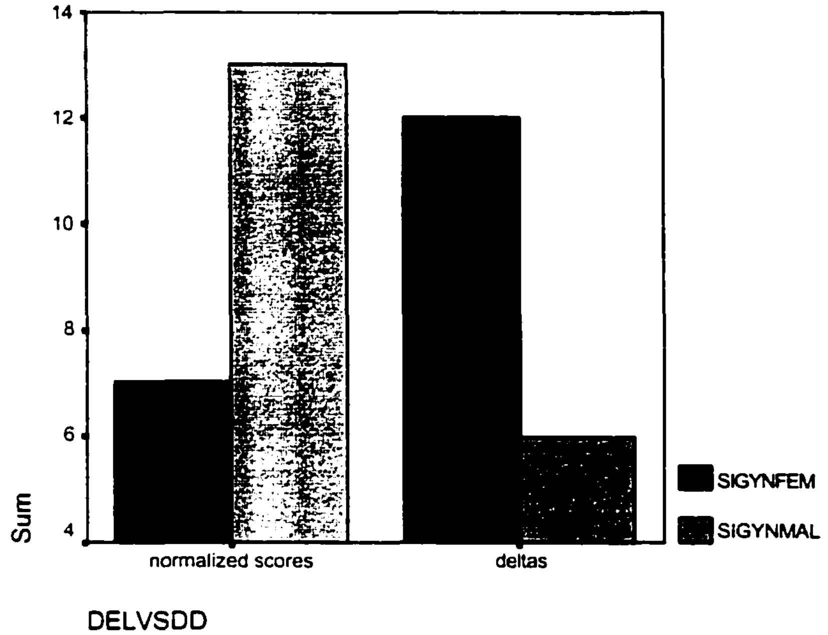


Figure 47. Number of significant correlations (0 = no correlation, 1 = a correlation, either positive or negative), by interlocutor gender and score vs. delta.

The other interaction involving interlocutor gender and was between it and the exact measurement (i.e., normalized scores and deltas and medians, SDs, and ranges). This interaction was only significant for positive and negative results grouped together: $F(5) = 2.319, p < 0.05$.

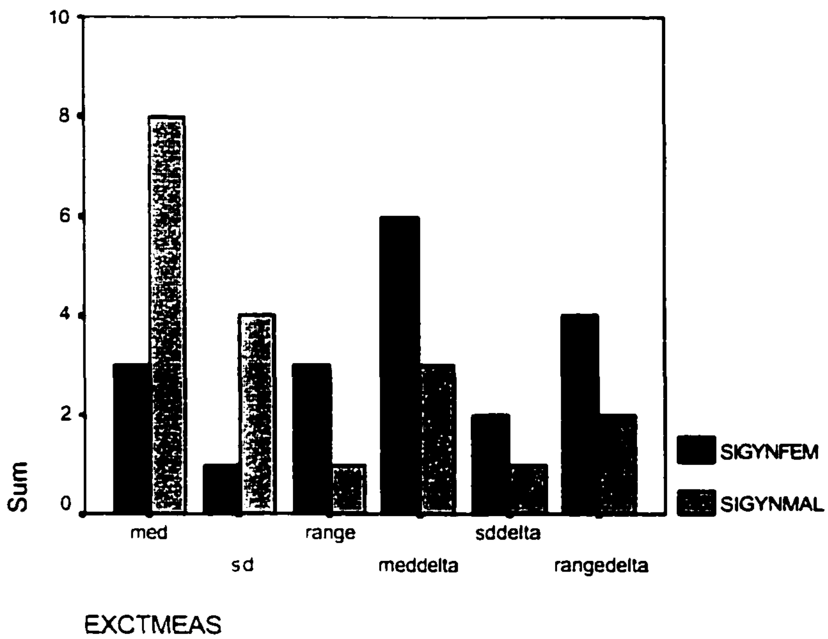
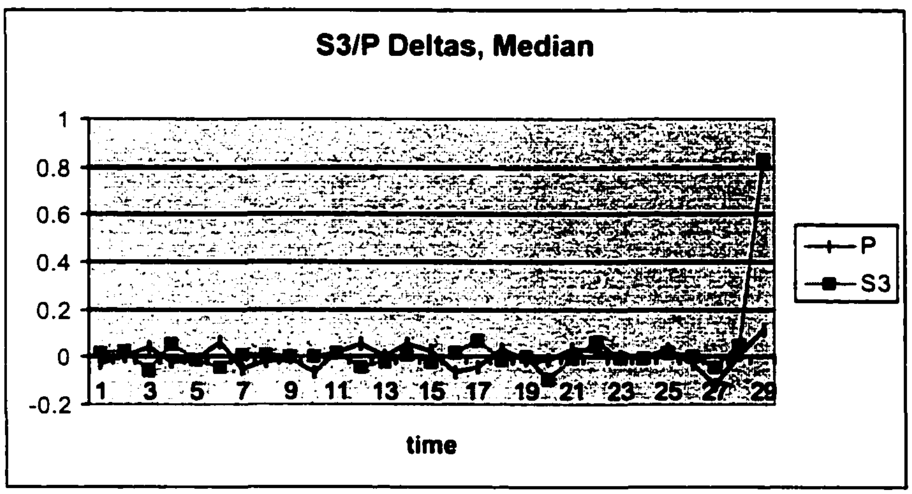


Figure 48. Number of significant correlations, by interlocutor gender and exact measurement.

This result suggests that the above normalized score vs. delta finding depends to a degree on whether the measure was median, SD, or pitch range. Female interlocutors had more delta-based results for all three measures, but males had more normalized score-based results only for median and SD; female interlocutors had more normalized score range results than males.

3.2.2.3. Suspicious Cases

The above descriptions are based on the data as measured numerically, with very little interpretation. If an individual correlation was significant at the 0.05 level, it was counted as a significant result. However, in the course of examining the scatterplots and line graphs to discern the relationship between the grouped and individual significant results, a few suspicious cases were found. Six of the significant individual correlations appear to be due to outliers, at least as they appear in scatterplots. They were: S1 with P, delta range; S3 with B, normalized score SD; S3 with P, delta median; S8 with J, normalized score, median; S8 with J, delta median, and S11 with J, normalized score range. Scatterplots and line graphs with and without the outlier for two representative cases, those for S3 with P, delta median, and for S8 with J, delta median, are below. Graphs for the other cases can be found in Appendix 6.



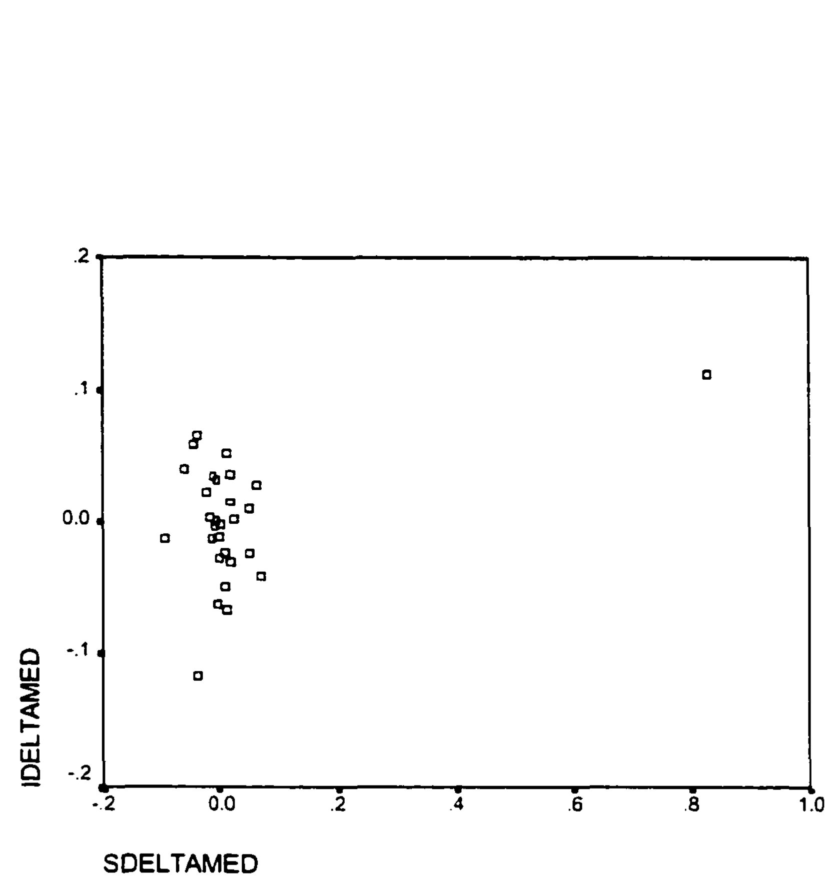
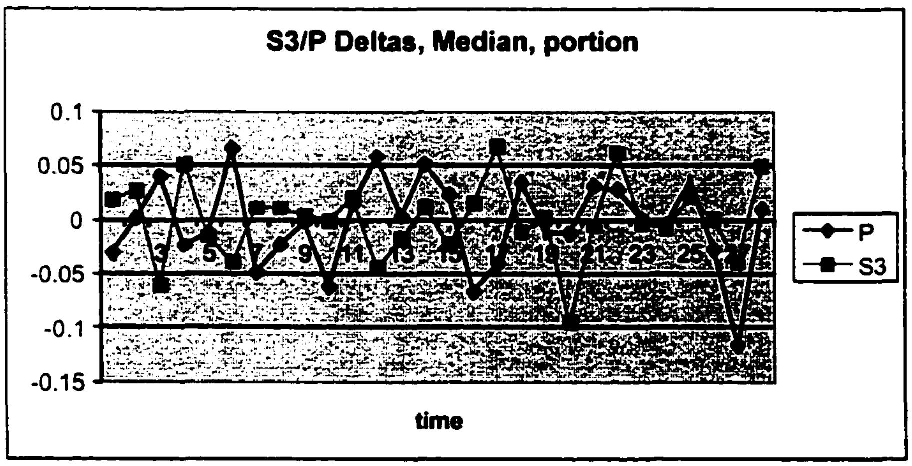


Figure 49. S3 with P, deltas, medians, with all data points.



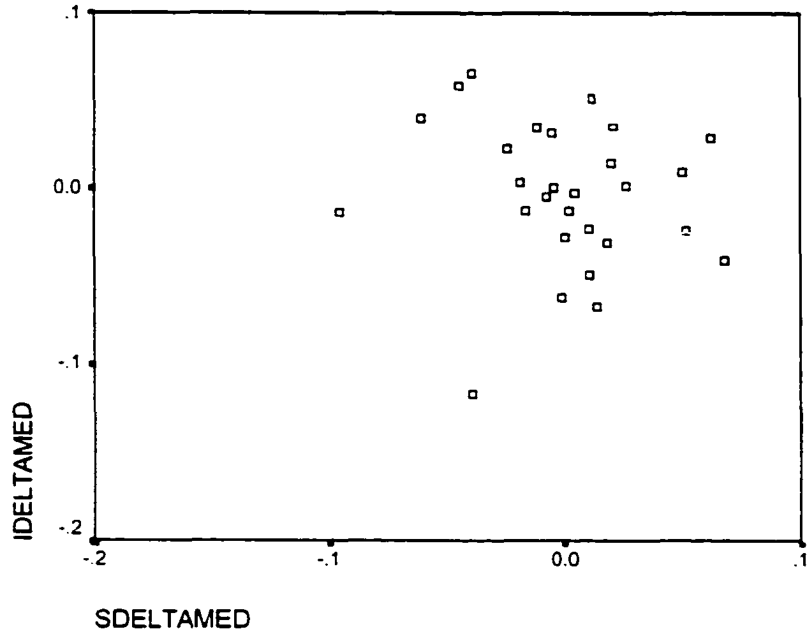
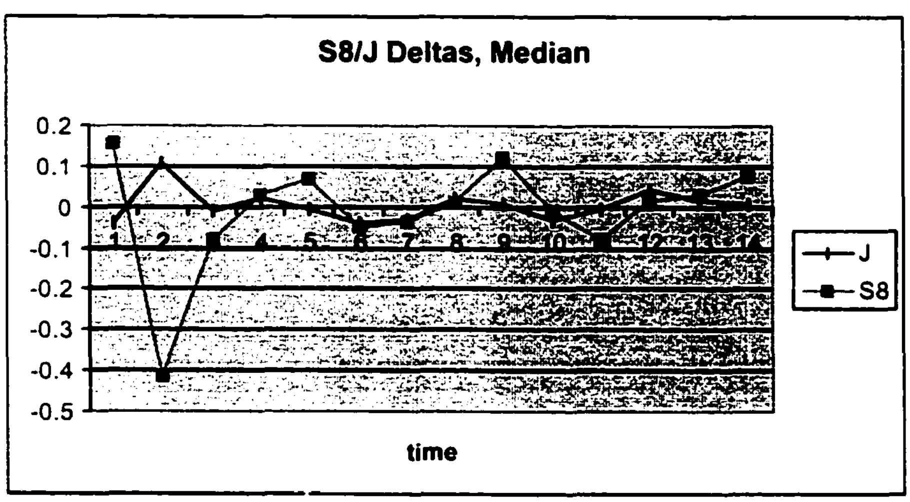


Figure 50. S3 with P, deltas, medians, without outlier.



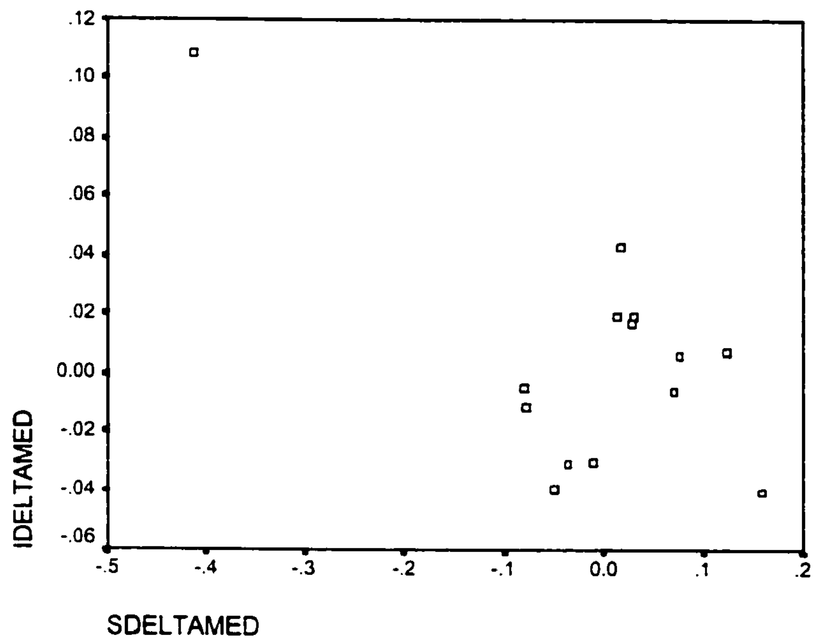
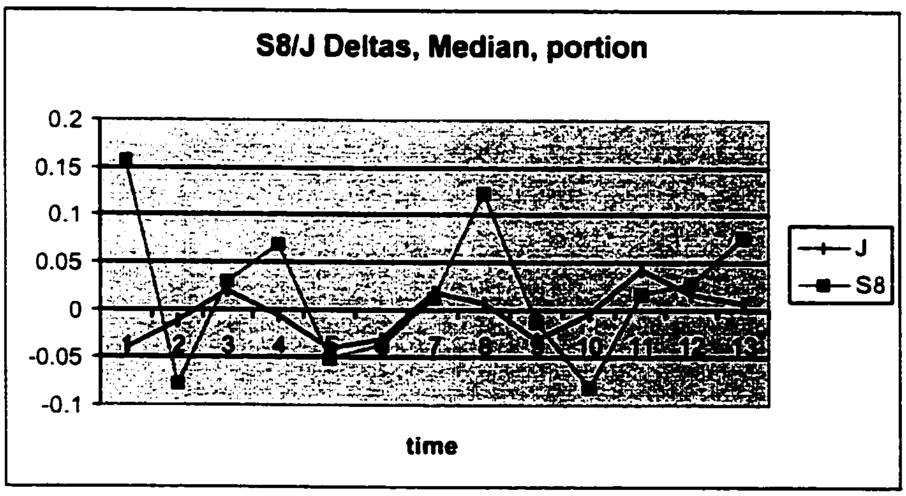


Figure 51. S8 with J, deltas, medians, with all data points.



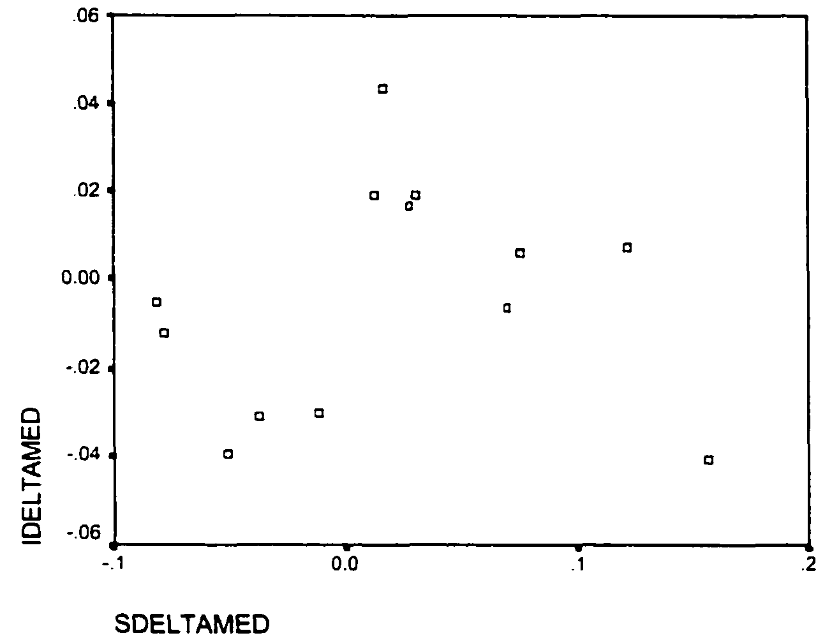


Figure 52. S8 with J, deltas, medians, without outlier.

There is other supporting evidence that these significant correlations were due to outliers. When the one suspicious data point (or in two cases, S1 with P, delta range, and S8 with J, normalized score median, two points) was removed, the Pearson correlation coefficient was dramatically reduced.

	ORIGINAL PEARSON COEFFICIENT (r)	COEFFICIENT W/O OUTLIER(S)
S1P	r = 0.590	r = - 0.0004
S3B	r = - 0.632	r = - 0.0001
S3P	r = 0.432	r = - 0.127
S8J, norm. sc.	r = - 0.542	r = 0.056
S8J, delta	r = - 0.623	r = 0.460
S11J	r = 0.551	r = 0.168

Table 11. Pearson correlation coefficients with and without outlier(s).

Furthermore, when I checked the amount of speech material in the time chunks responsible for these suspicious data points, for six out of the seven relevant time chunks (one for each of four cases and two for two cases, but since the interview S8 with J is involved in two of the suspicious cases, one time chunk is involved twice), the speech material measured was less than five seconds in total. The evidence points to these outlier data points as being spurious measurements.

However strikingly suspicious these cases are, my interpretations are just that, (subjective) interpretations. Furthermore, they do not affect the above patterns greatly. There are three normalized score results and three delta results that could be disregarded, and three median, one SD, and two range results, so the categories would all be relatively equally affected if these cases were not included. Negative results would be affected most, as three out of six suspicious cases are negative, but the patterns discussed above would not be changed: the adjusted interlocutor groups numbers would be four for males versus one for females, and four for professors versus one for peers. It is interesting to note that two of the subjects would have no correlations if these cases are not considered, S3 and S8.

The significant correlations discussed in the above sections are being considered evidence of pitch accommodation. The broader interpretations and ramifications of the patterns among the subjects and interlocutor groups found for both experiment one and experiment two will be explored in Chapter 4.

Chapter 4. Discussion

4.1. Explanations for the Patterns Found

Experiment one, based on the female subjects' pitch measurements grouped according to the gender and status of their interlocutors, revealed one significant trend in the data pooled from all 12 subjects; pitch range was larger with female interlocutors compared with male interlocutors. This experiment also found two groups within the twelve subjects, a group whose pitches varied according to interlocutor gender, and a group whose pitches varied according to interlocutor status. The gender group (S4, S8, S9, and S12) had more varied pitches, either a larger standard deviation of pitch or a larger pitch range, with females, echoing the trend in the pooled data. Two subjects (S1 and S10) had higher median pitches with females and could be considered a part of this group, as well. These more varied and in some cases higher pitches may be reflective of a high-engagement interaction style, i.e., a solidarity-based, friendly politeness style. The status group had two contingents, two subjects (S6 and S7) who had higher median pitches with professors, and two subjects (S2 and S5) who had more varied pitches with peers. The higher median pitches with professors may be reflective of a deference politeness style, while the more varied pitches with peers may again point to solidarity-based politeness.

Experiment two, based on pitch measurements from both subjects and their interlocutors to measure any accommodation in the pitches used within interviews, revealed several significant findings, but the subjects did not cluster together as tidily. Firstly, it is possible that some of the accommodation findings explain peculiar results from experiment one. The one subject who had a gender-based result that did not match the others' pattern was S2, who had higher median pitches with males. She showed evidence of accommodation in her interview with the male professor for median pitch (normalized scores). The two other slightly different gender-based subjects, S1 and S10, who had higher medians with females rather than more variability, also showed some evidence of accommodation in that they each had a significant correlation in one of their interviews for median pitch, S1 with the male professor and S10 with the female professor. These results were for deltas, not normalized scores, however. Thus, as mentioned in Chapter 3, the conflicting cases in the first experiment all had some evidence of accommodation in experiment two. But, it is important to recall that a few of the subjects that patterned together neatly in experiment one (S5, S7 and S12) had some evidence of accommodation in the relevant measures, as well.

As was also discussed in Chapter 3, there were a few patterns based on the social characteristics of the interlocutors that particular subjects accommodated with/to. S5 only accommodated with females, and S9 mostly did (four significant results with females and one with males). S12 may also be considered to be in this group of subjects, as she had five results with females and two with males. S10 had all of her five results with professors, but it is noteworthy that four of them were with the female professor and only one with the male professor. She may be considered as a part of the "accommodating with females" group, as well, in that she did have four results with a female and one with a male. Given this interpretation, three (S9, S12, and S10) of the gender-based subjects who had higher or more varied pitches with females in experiment one had more accommodation with females in experiment two.

My original aim was to use the two experiments to isolate the two factors that could be responsible for socially-based, intra-speaker pitch variation, i.e., expectation-based variations and accommodation. They are indeed distinct phenomena, and each encompasses an array of strategies available to speakers. However, upon closer inspection, they may not be motivated by separate phenomena, and it may not be an accident that there is some overlap between the gender (female)-based subjects in the two experiments. When the literature on positive politeness strategies, which I am suggesting are responsible for the gender group's pitch patterns in experiment one, is compared with that on accommodation, there are parallels. Lakoff (1973) defines three rules of politeness that are part of speakers' pragmatic competence. The first two, "Don't impose" and "Give options," protect the addressee's negative face, but rule three, "Make A feel good – be friendly," involves positive politeness in that it invokes a sense of camaraderie between the speaker and interlocutor (p. 298 ff.). Utterances involving rule three aim to produce a sense of equality and friendliness—solidarity. Brown and Levinson (1987) developed, formalized, and explored these two kinds of politeness strategies, invoking Durkheim's (1915) and Goffman's (1967) notions of positive and negative face. Claiming that individuals have both positive and negative faces they wish to protect, they define face as individuals' self-esteem (p. 2), with positive face pertaining to an individual's desire that his or her wants be thought of as desirable (p. 101) and negative face pertaining to an individual's wants for freedom and to not be impeded (p. 129). They then describe in detail different politeness strategies that provide redressive action to an interlocutor's positive or negative face when an utterance could be taken as offensive to it.

Brown and Levinson have been criticized on many fronts and the universality of their claims questioned (see Kasper, 1990, Meier, 1995, and Jary, 1998 for examples and more critical references), but the general ideas of positive and negative face and politeness strategies have proven useful in a large body of research, and other researchers have defended the utility of Brown and Levinson's claims, especially if they are adjusted slightly. For instance, Chen (2001) works within their framework but suggests that the speaker's face needs as well as the addressee's face needs must be considered when accounting for behavior, and his ideas intuitively seem correct. In a sense, however, the speaker's and addressee's face needs are two sides of the same coin in that speakers are able to anticipate addressees' face needs to some extent because they have the same kinds of needs (although perhaps in different proportions). Paying attention to the addressee's face needs could be considered an exercise in empathy. It could furthermore be seen as evidence of the speaker's positive face, regardless of whether the addressee's positive or negative face needs are being considered, as a speaker's empathetic actions communicate the desire to be seen as nice, considerate, etc. Thus, I am considering positive and negative face needs and strategies in a general sense, as I believe the speaker's and interlocutor's needs are connected.

Tannen (1984) uses these face needs, described as competing needs for closeness and independence, to account for two kinds of speech styles she encountered in her collection of data among friends and acquaintances at a Thanksgiving dinner. The style based on positive politeness strategies, the high-involvement style, places the most importance on rapport and signaling that one is engaged in the conversation. One of the features of this style is marked pitch and amplitude shifts (p. 31); marked pitch shifts may have caused the larger standard deviation of pitch and pitch ranges found in some of the conversations between female interlocutors. Tannen's description of high-involvement style lends support to an interpretation of the gender-group's behavior in experiment one as evidence of a camaraderie-based politeness or conversational style.

As alluded to above, positive politeness and accommodation may not be unrelated. Positive face needs are about wanting to be liked and wanting to fit in. The April 1, 2002, issue of *Time* magazine ran a story on the increase in binge drinking among young women. A female college student was quoted as saying, "When you drink, guys think you're cool...of course, the definition of cool is to be more like themselves" (p. 6). When researchers such as Bell (1984) have tried to formalize this intuition, they have drawn heavily on the work of social psychologists, as the motivations for accommodation undoubtedly rest in this realm. In their 1975 book, Giles and Powesland articulate a model of accommodation stemming from similarity-attraction research that suggests that "an individual can induce another to evaluate him more favorably by reducing dissimilarities between them" (p. 157). It assumes that the desire for social approval is at the heart of accommodation and that social approval is a reward in and of itself (pp. 158-159).

Within their model, some speech characteristics are subject to "covert" accommodation "of which the speaker himself may have little or no consciousness, [and] would probably include changes in speech rate, pauses, grammatical complexity, accent etc. made by way of convergence towards the speech characteristics of the receiver" (pp. 169-170). It is my contention that pitch characteristics fall into this category, unlike shifts between languages within bilinguals, the case that is discussed most thoroughly in the book. Another distinction between such overt and covert accommodation is that it is much easier to tease apart who is accommodating in overt accommodation situations such as when one bilingual switches to another bilingual's dominant language. When pitch characteristics are the variable being studied, it is much more difficult to discern who is accommodating or if both participants are. One possible hint provided by Giles and Powesland's model is that, in their description of the distinction between what they call asymmetric and symmetric convergence, they suggest that symmetric (i.e., mutual) convergence may occur "when no particular outcome of the interaction is expected, apart from the social support and satisfaction provided by the relationship itself…[i]n fact any instance in which two people are 'thrown together' by circumstances in a strange or hostile social environment would be conducive to mutual convergence for the sake of mutual solidarity and support" (pp. 176-177). The situation created by the experimental design might fall into this category of social situations, although it was not set up as hostile in any way.

This possibly mutual accommodation is thus explicitly tied to solidarity. It seems to be the case that the camaraderie-based (according to my interpretation) increased pitch variability with female interlocutors found in the gender-based subjects in experiment one and the accommodation found in experiment two are both linked to positive face needs. Both can be seen as expressions of solidarity that communicate "we're the same and we like each other" to the interlocutor. They both acknowledge the participants' desire for solidarity.

Accommodation in the realm of intonation and pitch has been addressed in some discourse analysis literature, without it being termed accommodation. Wennerstrom

(2001) defines "key" as the "relative placement of the pitch in a speaker's range at the onset of an utterance, which conveys the speaker's stance with respect to the prior utterance" (p. 273), and cites Brazil's (1985) description of "tone concord." Tone concord occurs when "one speaker's key matches the previous speaker's pitch termination" and indicates a supportive situation where the speaker and interlocutor are in agreement in their assessment of the conversational context (Wennerstrom, p. 180). It seems that such tone concord would likely produce pitch accommodation as I have measured it, i.e., pitch correlations. Wennerstrom asserts that metaphors for a successful conversation such as "in tune" and "harmonious," as well as "in synch" and others, indicate that "at some level, speakers of English associate a smoothly running conversation with pleasant, rhythmic music" (p. 261). A good conversation is when speakers are "on the same wavelength" (*ibid*). The same wavelength suggests a situation with camaraderie and agreement—solidarity and positive politeness.

It is interesting to note that Lakoff's (1973) research described above asserts that the positive politeness rule, R3, is basic and takes precedence over the other two rules that invoke distance or deference in that if it is expected, the negative politeness strategies will not be interpreted as such but will be considered rude. Furthermore, the extent of positive politeness strategies in my data could be seen as supporting evidence for her claim that R3 is becoming more and more prevalent at the expense of the other strategies in middle-class America (p. 302).

There were also two subjects (S4 and S11) who had all of their significant results in experiment two with males. It may well be that some individuals (including some females) use positive politeness strategies with males more than they do with females, at least in certain situations.

While strategies that reflect the importance of positive face needs were prevalent, there were some findings that point to the presence of negative face needs. Aside from the two subjects who had higher median pitches with professors in experiment one (S6 and S7), there were several instances of negative correlations in experiment two. They were (not considering the suspicious cases that were probably due to outliers): S4, SD with J, S11, median with J, S11, delta median with J, S12, median with J, and S12, delta median with A. (The suspicious negative cases were S3, SD with B; S8, median with J; and S8, delta median with J.) There are several possible explanations for the divergence or dissimilation of interlocutors' pitches in a conversation. Bell (1984) discusses referee design where a speaker assimilates to a group that the addressee does not belong to; this pattern might occur when a speaker wishes to heighten signals of belonging to a group that the addressee does not belong to, for example, when a female is talking to a male. This explanation is problematic, however, as the negative correlations happen more with the male professor than with the male peer, which would be hard to explain if the female speakers were heightening female group membership cues (to heighten the contrast with males) via higher pitches. Furthermore, in experiment one, the higher pitches occurred with professors as a group, rather than with males as a group.

A more likely explanation of the negative correlations is an extension of the deference politeness strategy offered as an explanation for subjects having higher median pitches with professors in experiment one. Negative correlations, particularly those for median pitch, may be due to increased symbolic deference in the face of perceived threat or symbolic authority. If certain subjects were using higher pitch out of deference, they may well have heightened their use of this strategy when their interlocutors lowered their pitch and invoked (intentionally or not) more authority. The association of low pitch with authority or dominance or threat and high pitch with deference or submission is tied to sound symbolism. Ohala (1994) proposes a "unifying, ethologically based and phonetically plausible theory of aspects of sound symbolism" (p. 326) that can explain sexual dimorphism in the vocal apparatus as well as social manipulations of pitch. Certain tendencies have been established: Languages use high and/or rising F0 for questions, deference, and lack of confidence and low and/or falling (especially terminal fall) F0 for statements, authority, aggression, and confidence. Other species have the same F0-meaning correlation, the "frequency code," that aggressive sounds are often rough and low and submissive ones are high and often tone-like. F0 can convey the size of the signaler, since it is inversely related to the mass of the vocal cords and overall body mass (Morton, 1977, p. 330). F0 is used to make the hearer react as if the signaler (or something in the environment) were small/large, making use of an innately determined frequency code. Signals are designed to produce favorable responses in the hearer. For example, submissive displays often involve high pitch which conveys an impression of smallness (as small creatures tend to make high-pitched noises) and lack of threat. Such noises are in fact rather like infant vocalizations, which could inhibit harming behaviors or stimulate helping responses. On the other hand, low-pitched vocalizations signal threat or confidence, conveying the image of largeness, which could be beneficial to the signaler in other circumstances (Ohala, 1996). Sexual dimorphism in vocal anatomy, i.e., a lowered and bigger larynx in males, develops to enhance aggressive displays acoustically in males by enabling them to make lower-frequency noises at the time (puberty) when they will be needed.

A seeming contradiction to an ethological basis for sexual dimorphism in the vocal tract and the frequency code is in the evidence that height apparently does not correlate with fundamental frequency (Künzel, 1989). The frequency code is based on the idea that low-frequency sounds reflect large creatures and high-frequency sounds reflect small creatures, and we are all aware of these connections and can count on others to be aware of them. What if low-frequency sounds do not only come from large creatures and high ones do not only come from small creatures? A response to this concern can be gleaned from an interpretation of Graddol and Swann's (1983) discussion of their finding that there is a correlation between F0 and height for men but not for women: "Since men's heights can actually correlate better with their median SFF [speaking fundamental frequency] than with their basal F0 [lowest possible pitch], it may be that men are adapting their SFF behaviour in some way in order to reflect their body build…to maximize *apparent* height, perhaps because height is an important part of the male gender image" (p. 363, my emphasis). F0 may reflect projected height more than actual height, within what is physically possible. But, this situation does not at all invalidate the frequency code. The frequency code is a code, that is, it pertains to the symbolic connection between signals and meanings. Although it is not arbitrary, it perhaps has evolved over time. Its evolutionary value (its phylogenetic adaptivity) is related to the fact that its users can manipulate it. A signaler can evoke the connection between, e.g., low pitch and largeness, and convey a larger size than the signaler actually possesses, evoking the desired response regardless of the signaler's actual size. Due to its manipulation by its users, the frequency code is perhaps now more of a symbolic representation of size rather than an actual representation of size.

The above ethologically-based deference politeness explanation makes most sense for higher median pitch and negative correlations for median pitch. But some subjects had negative correlations for standard deviation of pitch. A possible interpretation of these results could be something parallel to Yuasa's (1998, 1999) analysis of Japanese speakers' narrow pitch ranges with unfamiliar interlocutors reflecting a distance politeness style based on restraint. Perhaps these subjects felt uncomfortable in certain interviews and their discomfort increased when their unfamiliar (and often professorial) interlocutor got more animated. This interpretation is speculative, however, and negative correlation also in theory means that the subject's pitch measure increased when the interlocutor's lessened. Perhaps an ethologically-based explanation is appropriate for this end of the range of measurements. However, while piecing together different strategies to explain different parts of the data may be appropriate, it also runs the risk of being *ad hoc*.

Thus, the two politeness strategies or styles offered as an explanation for the results from experiment one, camaraderie and deference, both appear to play roles in explaining results from experiment two. Appeals to speakers' and addressees' positive face via solidarity may lie behind the more variable pitches found in experiment one with female interlocutors as well as the large number of positive correlations taken to be evidence of accommodation found between subjects' and interlocutors' pitch measures in experiment two. Appeals to negative face via deference, on the other hand, may explain the data from the two subjects who had higher median pitches with professor interlocutors in experiment one as well as (although this part of the analysis is more tentative) the small number of negative correlations found in interviews in experiment two. It is striking that the instances of pitch patterns consistent with positive politeness strategies far outnumber those consistent with negative politeness strategies. As was mentioned above, this ample evidence of solidarity-based politeness strategies in these female subjects may point to the current importance of solidarity in white middle-class culture, perhaps especially among females. It may be the strategy of choice among many female speakers of at least California English.

There are other patterns in experiment two that need explanation. Many of the correlations (both positive and negative) were found for median pitch rather than for standard deviation of pitch or pitch range. Why does more accommodation occur for median pitch? Another rather robust pattern that needs explanation is that accommodation occurs more for delta measures with female interlocutors but more for normalized scores with male interlocutors. The delta measures with females may be tied to a camaraderie strategy, but this interpretation is extremely speculative.

4.2. Previous Work on Politeness with Different Interlocutors

There has been some previous research that focused on politeness strategies varying according to social characteristics of the interlocutor. They have measured several different variables that have been tied to politeness and have interpreted the patterns found. One study, Brouwer (1982), even found a small amount of evidence of accommodation. This study replicated the results of Brouwer et al. (1979) discussed in Chapter 1 that travelers purchasing tickets in a Dutch train station were more polite to male ticket sellers than they were to females. However, many of the politeness markers studied were negative politeness strategies (e.g., civilities, hesitations). There was also a slight tendency for more diminutives in same-sex interactions. Diminutives are described as markers of nonchalance (p. 700) more than markers of endearment, although they can also serve this function. More diminutives in both female-female and male-male conversations may indicate that the subjects were more comfortable and perhaps felt more camaraderie in same-sex conversations. Interestingly, one female ticket seller (out of five) showed accommodation in the amount of politeness she used; the ticket seller that travelers were most polite to was the one who used the largest number of politeness markers that were studied. Brouwer interprets this finding: "It seems that polite behavior on the part of the travelers made less impression upon the male than on the female clerks since they showed a greater adaptation" (p. 708). The finding that a female clerk accommodated to the level of politeness she encountered is noteworthy and should be pursued, but the pattern was not robust enough to draw general conclusions.

While the above study focused mostly on negative politeness strategies, several other studies compared kinds of politeness used with different interlocutors and looked specifically at positive politeness strategies. Johnstone et al. (1992) had an experimental design that complements mine. They explored the effects of interviewee sex on interviewers' speech, manipulating social characteristics of the interviewee rather than the interviewer, as my study did. Their data consisted of taped public opinion surveys conducted over the telephone, and all of the interviewers were female. The interviewers adjusted their conversational strategies according to the gender of their interlocutor in order to do their job of keeping conversations going and obtaining information in the format they needed it. In addition to needing to manage the conversations with males more than those with females, the interviewers used different politeness strategies with males and females. While the numbers of instances of both positive and negative politeness strategies were similar with female and male interviewees, different strategies were predominant with each group. For positive politeness strategies, female interviewees elicited more demonstrations of sympathy and understanding, but male interviewees elicited more attention to their wants and needs (including laughter) and more joking. For negative politeness, apologies occurred slightly differently with males and females. More interviews with females had an apology in them, but when a male was apologized to, it was much more profuse and repeated. Females were also thanked more.

Johnstone et al. interpret the above patterns as evidence of the ability of the female interviewers to communicate via audience design and adapt their strategies according to what will work with a particular interlocutor. The interviewers both expect and react to males' apparent discomfort at being in a somewhat powerless position as the conversation is necessarily run by the interviewer. They subtly corral them in the direction needed and facilitate their responses as much as possible. They respond and even accommodate to the joking used by the males at times to derail the interview, continuing joke exchanges initiated by males before moving on. When the males were apologized to if it was suggested that they were being inconvenienced, they often ignored or dismissed the apology and it was repeated more profusely. According to Johnstone et al., the female interviewees were more cooperative and thus needed less discourse management and were appropriately thanked more. When they were apologized to, they accepted or at least recognized the apologies.

The positive politeness strategies used with female interviewees were those based in solidarity, i.e., those emphasizing mutual understanding of the other's position. Connor-Linton (1986) studied politeness strategies in white middle-class adolescents in Southern California and also found evidence of the importance of positive face needs among female speaker-listeners. In his recorded conversations, he found that while males and females used approximately the same amount of politeness, there were some differences in the strategies used, and males were the preferred targets for turns, politeness, as well as face-threatening acts. Connor-Linton suggests that females focus much of their negotiation energy towards males in attempts to obtain some control in the conversations. While they directed their bids for control towards males, "no female committed an FTA against another female's positive face" in the data collected (p. 81). Despite varied and strategic use of different maneuvers throughout the conversations, apparently the females' positive face needs were completely respected within females. This finding again highlights the importance of solidarity and positive face in general within the females involved in these studies. Sheldon (1992), studying much younger children (ages three to five) in daycare, also found that the girls were not at all passive in negotiating their needs in their actions. Furthermore, she found that they utilized what has been called double-voice discourse in which they pursue their own agendas while simultaneously orienting towards their interlocutor's needs and points of view. The constant orientation towards others' face needs also points to the importance of empathy, sympathy and solidarity among even very young females.

The recurrent theme in many of these studies is the central role that solidarity or camaraderie plays in female-female interactions. The fact that the strongest patterns in both of my experiments with female speakers point to positive politeness strategies is in line with this idea. The more variable pitches found with female interlocutors in experiment one and the accommodation found with both female and male interlocutors in experiment two suggest that pitch patterns are yet another way that solidarity politeness strategies are expressed in conversations. There is evidence of deference politeness in the pitch data, as well, but it seems to play a much smaller role. As was mentioned above, Lakoff's (1973) assessment/prediction regarding the increasing predominance of positive politeness strategies appears to be borne out in my pitch data as well as in several other studies on politeness strategies varying according to the gender of the speaker and interlocutor.

4.3. Perceptions and Evaluations of Female and Male Voices

The above discussion interprets the patterns that were found in two experiments measuring speakers' pitches and situates them within some of the previous research. It attempts to describe what the speakers' intentions might be, i.e., what they might have been intending to convey. However, what speakers intend to signal may not match what listeners perceive. Interlocutors' evaluations and interpretations of speakers' pitch patterns may or may not be in line with what speakers are attempting to communicate. A significant amount of research has been conducted on how some pitch patterns are perceived and on how women's voices can be evaluated; it will be reviewed below.

Van Bezooijen (1995) suggests that culturally-motivated differences in pitch (see Chapter 1 for a discussion of some of these) have their roots in culturally-based gender role expectations and sought to test whether reactions to women's voices would correlate with such expectations. Samples of Dutch and Japanese women's voices were manipulated so that there were original, higher pitch, and lower pitch versions of samples from each culture. The samples were of read, not spontaneous speech. Dutch and Japanese listeners rated the voices. For listeners (male and female) from both cultures, the ratings for the low pitch versions were higher than those for the high pitch versions for the characteristics *tall, strong, independent*, and *arrogant*. This supports Ohala's frequency code idea since higher pitch leads to a perception of *short, weak, dependent,* and *modest* in both cultures. The manipulated versions were also considered less attractive than the natural samples. Culture affected attractiveness ratings of the samples, however: Original and lower pitches were less attractive for the Japanese voices. Also, Dutch listeners considered low and original pitch more attractive than Japanese listeners did, and Japanese listeners considered high pitch more attractive than Dutch listeners did. These results show that what is attractive is culture specific. Furthermore, the cultures describe ideal men and women differently in a way that is consistent with the ratings, too. The differences between the ideal man and woman were bigger for Japanese subjects than for Dutch subjects. Dutch subjects rated the ideal woman as more independent than Japanese subjects did. Also, women in both cultures rated the ideal woman as more independent than men did. Thus, since there is evidence for the frequency code across cultures and stronger differentiation between ideal men and women in Japanese and the preferred pitch for women is higher in Japanese, it is plausible that Japanese women raise their pitch to conform to sociocultural expectations stressing femininity (p. 264). What surprised the author was that the ideal man in the two cultures differs more than the ideal woman. There seems to be a strong emphasis on masculinity in men in Japan, which could lead to pitch lowering.

Van Bezooijen (1996) continues this line of research of finding evidence for the influence of sociocultural factors on the mean pitch of female speakers from a perceptual viewpoint. It explores the extent to which impressions of gender-related personality traits are affected by pitch differences both within and between speakers, whether sex of the listener is a factor, and whether any correlations vary according to the language spoken by the speaker (p. 758). The speakers in the study were eight Dutch, eight Belgian (Flemish speaking), and eight Japanese women (ages were 20s-40s) who read a neutral narrative text. The speech samples were manipulated to create lowered, original, and raised pitch versions. Fifteen male and 15 female college students (Dutch) were listener-subjects who were asked to rate the speech samples on seven-point bipolar scales about personality traits. They were then asked to rate what they saw as the ideal man and the ideal woman on the same scales. Listeners of both sexes agreed about the ideal man and woman ratings. The ideal woman is seen as less independent, less arrogant, less insensitive, less rational, and having lower prestige than the ideal man. ANOVAs were done to see if these five gender-related factors were attributed differently to the same speakers with different pitch levels. Sex of listener did not matter, and language of speaker only did slightly. (The *listeners* were all Dutch, though.) Pitch was a significant factor: When the female voices were presented with a relatively low pitch, the speakers were perceived as more independent, arrogant, prestigious (higher prestige), insensitive, and rational, i.e., more masculine, than when they were presented with a relatively high pitch (p. 762). The nonmanipulated voices' pitch and personality scale ratings were compared, too—these correlations were also significant and in the same direction, but these natural voices varied along more parameters than pitch. The author suggests that the findings reflect the fact that women exploit the correlations between pitch and attributes (citing Ohala's frequency code) and adapt their voices accordingly to convey "the desired degree of femininity" (p. 765). The findings certainly reflect stereotypes about how women (should) talk and show how they are evaluated, but they may or may not reflect what the women are actually doing purposefully.

Rather than studying pitch's influence on perceived personality traits, Linke (1973) explores its effect on "effectiveness." Linke cites several studies that disagree with each other regarding women's average F0 and some on what F0s correlate with vocal effectiveness, e.g., Snidecor (1940) reports a lower average pitch in "superior" female readers, but other studies have not found this correlation in males. High pitch variability (standard deviation of pitch) has been found to correlate with effectiveness in men but not women, possibly since effective women use low pitch, which correlates with low variability. This study's objectives were to get data on pitch level and variability from females representing all levels of ability (effectiveness ratings) and to find which attributes contribute most to judgments of general effectiveness in women. Samples were read performances by 27 female speakers of American English (college age), chosen so that their effectiveness ratings ranged from 0.83 to 8.44 on a nine-point scale. The median frequency for the speakers was 201 Hz. It was believed that the SD of 1.52 tones was "inflated," due to very low downward dips in inflection that went lower than sustainable pitches; the SD based on frequencies above the lowest sustainable pitch was 1.21 tones. The only good positive correlation found was between this recalculated SD and effectiveness ratings. Pitch level in general did not correlate with effectiveness, despite previous findings that low pitch was more effective. Because the non-reduced (i.e., as originally calculated) SD did not correlate, the author asserts that excessive use of low frequencies reduces general effectiveness. To test this idea, the proportion of frequencies below the sustainable pitch level used was computed, and it correlated negatively with general effectiveness. Thus, women appear to use lower pitch than was previously thought, even lower than would be "advisable" for "effective" speech. Because the median pitch used is low, options seem to be to use little variability or dip down "lower than is advisable" (p. 185). The author concludes that women are lowering their pitch excessively due to pressures telling them to do so (previous advice, low pitch of women in the media, etc.) and this strategy isn't leading to improved effectiveness and could lead to lessened expressiveness due to reduced variability or even laryngeal injury. The overall picture he presents is that low pitch is not well-received in female speakers.

The above studies together show that high pitch is seen as feminine, and thus desirable, in women, but it is also associated with negative attributes such as weakness, irrationality, etc. Part of this contradiction may be due to how femininity is defined in various cultures, i.e., what the ideal woman is considered to be like. Women may be encouraged to strive to embody such ideals in order to be feminine and accepted, but by doing so, they come to be seen as weak, irrational, etc. (cf. the double bind discussed in Lakoff, 1975), in part due to their phonetic strategies (and also via their biologically-based phonetic traits). The finding that women may be using a lower pitch than is effective may be reflective of their reacting to explicit instructions from society to use a lower pitch to be taken seriously. While the influence of the frequency code cannot be underestimated, this advice may also be indicative of a misogynistic view of higher pitch being bad, it being bad simply because it connotes femaleness, which is inherently negative. Three examples of such explicit instructions are provided below.

Etiquette books and speech improvement books alike express very strong antipathy towards high-pitched voices. They are very unabashed in their disapproval and very direct in their recommendations. In a section called "A Modulated Voice," *Vogue's Book of Etiquette and Manners* (1969) discusses the following quote from Shakespeare's *King Lear*: "Her voice was ever soft,/Gentle, and low, an excellent thing in woman" (Act IV, scene 3, line 274). While "low" in this quote could pertain to either loudness or pitch or to both of them, the commentary is as follows: "These lines by Shakespeare describe the ideal feminine speaking voice; masculine voices may not be gentle or soft, but they, too, should be modulated. Nothing can be more uncomfortable to the ear than high-pitched, loud, or harsh vocal tones" (p. 15). The ideas expressed are that women must speak gently and softly (men do not have to) and high-pitch is intrinsically bad, and is in fact on a par with harshness. In the chapter on public speaking, there is a section called "Using One's Voice" that asserts:

> The natural pitch of a voice determines, to a great extent, how audible it will be. Many low-pitched, resonant voices are pleasant and soothing to listen to, but they can be difficult to hear. Men with very deep voices should direct their speech to the last row in the audience, to ensure that everyone present will hear what they say. Women's voices, with their naturally higher pitch, frequently are easier to understand. One of the hazards for women speaking in public, however, is that tension can constrict the vocal cords, making the voice so shrill and high-pitched that it grates on listeners' nerves. The only cure for this is practice and enough experience in speaking to avoid nervousness (p. 532).

Again, the assumption seems to be that the quality "high-pitched" is unpleasant, although in this case the assessment is related to both pitch and voice quality in the context of nervous tension.

Another etiquette book, *Good Housekeeping's Book of Today's Etiquette* (Raymond, 1965), is remarkably consistent with the above sentiments in its views and advice. Chapter 2, "Be a Woman with Charm," was the most explicit in its negative view of high-pitched voices. It asserts that charming women have three elements of poise: the quality of repose, the habit of attentive listening, and an attractive speaking voice. "Some fortunate individuals are blessed by nature with a voice that is low, musical, and resonant, but most women must devote some conscious effort to developing good speaking habits. Remember to speak distinctly and in a tone low enough to block any tendency to be shrill. Many of us ought to lower our voices at least one tone and practice this until it becomes second nature" (p. 6). According to this book about how women ought to be, most women's natural voices are unacceptable so they should actively change them and change them to be lower pitched.

The value judgment that high voices/high pitch in general is bad—unpleasant and requiring change lest it offend someone—is echoed also in Jacobi's (1996) book on improving vocal techniques aimed at business people who feel that their speaking voices are holding them back and making them less effective. It discusses how to modify your voice and why you should and includes sections on voice quality, within which the author discusses pitch. The ideas in this book are not based on any empirical evidence and cannot be considered scientific in any way. However, they provide more evidence on what kinds of voices are assumed to be bad and what 'experts' suggest are good voices in American society.

In the introductory section on how the book will benefit you, the reader is encouraged to join others who have benefited from it, including "the senior account executive...who lowered the pitch of his voice to gain vocal resonance and strength—and credibility" (p. xviii). Chapter 1, "Improve your Business Voice," includes a section called "How a soft, breathy voice can lose a deal." There is a vignette about a businesswoman's presentation that failed because "her soft, breathy voice made her presentation sound weak and lackluster...she didn't get the account." (p. 2). Another example involves a situation "when a store owner's shrill, high-pitched voice annoys customers so much that they go elsewhere, even at higher prices..." (p. 4). Such examples reflect the apparent "given" that low pitch is inherently credible and high pitch and breathiness are not. The author provides exercises to "build a stronger voice" that involve breathing, vocal support, and posture. A humming exercise in which you learn to feel resonance in your chest as you use a lower pitch mentions that "speaking with chest resonance adds depth and authority to your voice" (p. 130).

After all the exercises to lower your pitch, there is an anecdote about a woman with a low voice being mistaken for a man: "Here was a woman who actually needed a *higher* voice! This was a very unusual situation; as a rule, women with deeper voices should welcome their lower tones. After all, a deep voice has a commanding presence. Just think of Kathleen Turner and Candice Bergen" (p. 135). Tips on avoiding strain are also offered, including not using an unnatural pitch: "In an effort to sound more authoritative, many speakers, especially women, try to lower the pitch of their voices. However, forcing the voice down unnaturally for any length of time strains the throat muscles and can make you hoarse. Even worse, a listener can usually sense when a person is forcing out an unnaturally low pitch ... Be aware that building resonance into a voice can make it sounds richer and deeper without actually being lower in pitch [the ultimate good thing]" (p. 153). It seems women are accused of raising their pitch for social reasons (not implausible) but they have to be concerned about falsely lowering their voices, too. But, since their voices in their natural state are probably "too high," they are in a double bind.

A misogyny-based and frequency-code influenced evaluation of high pitch may come into the reactions to high pitch in men's voices, as well. Higher-pitched male voices are also rather negatively evaluated, as the following two studies show. It may be bad for a man to use higher pitch because it is associated with femaleness, which is inherently negative (cf. men insulting each other by calling each other girls or women). It may also be the case that it is reflective of the desire to heighten the differences between men and women, rather than being directly related to a negative opinion of women.

Studies of such reactions to male voices include Brown et al. (1974) and Apple et al. (1979). In Brown et al., the speakers were two college teachers who spoke the sentence "We were away a year ago." The two voices were then resynthesized in 27 different ways, with all possible combinations of three levels each of rate, mean F0, and variance of F0. Thirty-seven male and female judges rated these altered voices on fifteen scales of adjectives with their paired opposites, which were grouped into two factors, benevolence and competence. Listener ratings were averaged. F0 variance manipulation's effects weren't completely clear, but decreased variance led to lower competence and benevolence ratings and increased variance improved benevolence ratings. Higher mean

F0 voices were rated less competent and slightly less benevolent. Thus, in this study, higher-pitched male voices were rated as more incompetent and more malevolent. Less varied (in pitch) speech was negatively evaluated, too.

Apple et al. had similar findings in their experiments, which focused on F0 and rate and their effects on several state and trait variables. Samples were naturally produced utterances with smaller manipulations from 27 speakers (male college students). Listeners were male and female college students. In experiment one, 20 (14 male, six female) listeners were asked to rate the truthfulness of the responses on a seven-point scale after being told that half the speakers had been told to lie and half to tell the truth. Low pitch enhanced credibility and the unmanipulated rate was judged the most credible. High pitch may have been seen as evidence of stress due to lying. The second experiment aimed to see if more enduring qualities would be ascribed if there weren't situational factors (e.g., the possibility of lying) to explain vocal parameters. Eleven listeners (nine male, two female) were asked to rate the same voices on nine bipolar adjective scales, focusing on both content and delivery. Results included higher pitch and faster rate correlating with decreased "potency." In a third experiment, judgments were again sought on speakers' affective state. Ten listeners (two male, eight female) were asked to rate the same voices on fluency, emphaticness, persuasiveness, nervousness, and seriousness on seven-point scales, paying attention to form and content. Nervousness was the only scale that had a significant main effect for pitch—it increased with higher pitch. Persuasiveness went down, but this effect was marginally significant. The authors conclude that acoustic properties of a message do affect judgments on various state and trait variables. Thus, in both Brown et al. and Apple et al.'s studies, high pitch and slow rate were judged negatively. In Apple et al., it was also noted that the pitch manipulations also affected pitch variance so the pitch effects could have been due to pitch variance effects, and in Brown et al., decreased variance was linked to speakers' being judged to be less competent and less benevolent.

It seems that, for both males and females, high pitch connotes a lack of competence. For males, the negative evaluation of high pitch is consistent (and low pitch is consistently evaluated positively), but the above survey demonstrates the multi-layered, rather ambivalent nature of reactions to women's voices. There are many negative reactions to women's voices in that their phonetic traits are often associated with undesirable qualities, e.g., high pitch and breathiness are associated with incompetence and a lack of authority. Furthermore, high-pitched voices are often deemed unattractive and unpleasant. However, there are positive reactions to these same traits, too, as in high-pitched voices being considered attractive and feminine. For pitch variability, there is less research available, but high variability can be positively evaluated for both males and females, being associated with effectiveness or competence for both genders. However, Henton (1989, 1995, see Chapter 1) reviews perceptions of women's voices being negatively evaluated for being "swoopy," this term being tied to possibly pitch range, dynamism, and/or variability.

4.4. Interpretations of Perceptions and Behaviors

4.4.1. Interpretations: Why Are Women Perceived in this Manner?

The ethological theory of the frequency code described above as an explanation of deference politeness strategies can also explain the negative evaluations of high-pitched voices. They innately connote helplessness and lack of threat. While these characteristics may not engender respect, they are not completely without value, as no one likes to feel threatened, and depending on the culture these qualities are not inconsistent with what is considered feminine to greater and lesser extents. The ambivalence towards low pitch in females may reflect the interplay between these factors, as well. Furthermore, it may be the case that low pitch in females is evaluated positively in that it is male-like and thus positively valued, but too much of a male characteristic in a female is perhaps threatening and is negatively evaluated (R. Lakoff, personal communication).

McConnell-Ginet (1983) also describes how the aspects of women's intonation patterns are associated with emotionality. Her analysis offers a possible explanation for the negative assessment of large pitch ranges and standard deviations of pitch, i.e., high variability. This connection between variability and emotionality could have come to be because there *are* physical causes of pitch changes during emotions; heightened breathing and muscular tension caused by strong emotion could lead to less pitch control—more dynamism and a bigger range (cf. Apple et al.'s, 1979, finding that higher pitch was associated with nervousness). This possible explanation for pitch variability might be behind the stereotype of women being emotional as variability that is not due to emotionality may be attributed to it. This characteristic is typically seen as bad, showing that the speaker is unstable, rather than expressive. However, the perception studies reviewed above show that high variability can be positively evaluated, as well. Further research on the presence of, reaction to, and causes of pitch variability is needed.

In fact, it may well be the case that high variability is positively evaluated in women as it is heard and encountered in actual speech, but at the same time it is consciously evaluated negatively. This and many other negative responses to and evaluations of female voices may have more to do with preexisting expectations and stereotypes listeners have about females and their voices. Even the consistent (and often positive) reactions found in listener-subjects in perceptual experiments may be influenced by stereotyping. Researchers have offered explanations of how gender-based stereotypes can come about and how they are not typically recognized as such.

In contrast to claims that gender is a social construction, there is a very strong social notion (at least within Western societies) that sex is unequivocally biologically based and clear-cut. West and Zimmerman (1991) suggest that gender itself is seen as natural and reflective of natural differences between men and women. It is composed of two natural categories that entail biological, psychological and behavioral traits (p. 15). Its distinctions reflect our ideas about inborn differences between the sexes and then serve to reinforce them: "Doing gender means creating differences between girls and boys and women and men, differences that are not natural, essential, or biological. Once the differences have been constructed, they are used to reinforce the 'essentialness' of gender" (p. 24).

Because we tend to see many traits as natural, we tend to trust our impressions of people. We react to them, e.g., to their voices, and form an impression and we assume that how they seem to us reflects not only their personality but who they are naturally. Listeners react to voices fairly consistently. People make reliable judgments based on voice samples in that they agree with each other and tend to form similar impressions based on a given voice. Many studies aim to see whether voice qualities correlate with personality traits, and some take consistent listener reactions to a voice as evidence.

However, what was being uncovered in these studies may well be the existence of stereotypes. (Many authors acknowledge this situation, and some even specifically strive to study stereotypes, e.g., Aronovitch, 1976.) Laver (1991) discusses this issue perspicaciously:

> We all act, as listeners, as if we were experts in using information in voice quality to reach conclusions about biological, psychological and social characteristics of speakers. Long experience of inferring such characteristics from voice quality, presumably often successfully confirmed by information from other levels, invests our implicit ideas about the correlations between voice quality and indexical information with an imagined infallibility. It is worth questioning the validity of this judgement process...There is a good deal of evidence that in such subjective judgements we operate with stereotypes...Listeners, if they are from the same culture, tend to reach the same indexical conclusions from the same evidence, but the conclusions themselves may, on occasion, bear no reliable relation to the real characteristics of the speaker (pp. 156-157).

Indeed, the question is how true are the stereotypes? Regardless of their validity, they play an important role in listeners' reactions to voices.

4.4.2. Interpretations: Why Do Women Speak as They Do?

In addition to the above interpretations of why listeners perceive voices as they do, scholars have offered interpretations for why women speak as they do, as well. As mentioned above and taken up again in Chapter 5, it is not clear why individual subjects in my experiments chose the strategies they chose. Ethnographic study may shed light on individual motivations. However, explanations for why women speak as they do as a group have been suggested.

Lakoff (1990) discusses social explanations for linguistic differences between men and women. She asserts that a nondominant group whose members don't directly have access to social, political, and economic power will have to find indirect ways to get their needs met. The strategies have acquired symbolic meaning, expressing distance from and lack of desire for power, and these strategies and even what they reflect have become part of what is deemed feminine and what makes a woman a nice person. Thus, there is a paradox for women in a powerful position requiring assertiveness and forcefulness today: "...she can be a good woman but a bad executive or professional; or vice versa. To do both is impossible" (p. 206). Despite this understanding of the development of women's language in English, Lakoff points out that the explicit behavior of disparaged and stereotyped women's language varies across cultures. However, regardless of the nature of the differences within a culture, the dominant group (men) will typically notice the difference between themselves and a nondominant group (women) and instead of attributing the difference to differing cultural expectations, etc., will figure it is due to some intrinsic physical or psychological difference between the sexes (cf. the discussion of stereotyping just above).

The phonetic differences between men's and women's voices could also be subject to this pattern of response, where differences between men's and women's voices, such as higher pitch in women, could be seen purely as a direct reflection of innate differences between the sexes. If there is an unequal distribution of power and the two groups involved are seen as inherently different, the differences could easily be seen as justifying the structural/hierarchical organization of a society, regardless of the nature of those differences. To take a non-linguistic example, the traditional division of labor allots childcare to women, and this work (at least in Western societies) is consistently devalued, and when it is paid, it commands a very low wage. This evaluation does not seem to be based on the importance of the work or its difficulty level (although in our society it does not entail as much formal training as some highly-paid kinds of work), but rather the work is devalued *because* it has been allotted to women. Some scholars claim the same basis of evaluation for aspects of women's voices. Cameron (1985) asserts: "It is inconceivable that these judgments [e.g., that women's voices lack authority] have anything to do with pitch. If men talked in higher pitches than women, low voices would be said to lack in authority" (Cameron, 1985, p. 54, cited in Graddol and Swann, 1989, p. 35). However, as Graddol and Swann point out, this interpretation suggests that the phonetic differences have no roots external to language and culture (or misogyny), and ignores the presence of "child voices, large growling dogs, and so on" in our surroundings (p. 36). As is discussed above, ethology can provide an explanation based in physical principles and cross-species behavioral patterns of why high pitch is evaluated in the way that it is that must be taken into account, as well.

In contrast to the interpretations that women's conversational and politeness styles came about and were reinforced by their lack of power, other scholars suggest that the different politeness styles used by women reflect their skill in conversation. In their analysis of the speech of females pollsters speaking differently with male and female interviewees, Johnstone et al. (1992), discussed above, recognize that the pollsters are reacting to the different levels of comfort that males and females have been socialized to feel in situations where they are in a less powerful position, but salute the pollsters' ability to adapt. They suggest that this job is female dominated due to females' "facility at varying their speech appropriately to elicit cooperation and information from men as well as from women" (p. 429). Sheldon (1992) and Holmes (1993) also point out the skill involved in and the positive results of females' attention to positive face needs. Holmes asserts that women are ideal conversational partners due to their sensitivity to others' face needs, regardless of their interlocutor's gender. She concludes her discussion of New Zealand males' and females' differential use of politeness markers and amounts of talk as follows:

> The greater use of face-supportive acts such as compliments and apologies between women suggests that women take positive steps to maintain verbal behaviours they value. They are not just passive respondents in interactions and nor is their behaviour determined by male norms. There are clearly definable female norms which put the addressee's interests and needs first. Consequently both women and men recognise that talking to women is a positive experience. But not everyone realises why (p. 113).

Adapting politeness strategies according to the interlocutor, accommodating in general, and keeping a vigilant watch over the interlocutor's positive face needs all undoubtedly require impressive and admirable skill. Both kinds of interpretations are valid at once; women's accommodating and positive politeness strategies are consistent with the language of the powerless and may reinforce powerlessness in some circumstances, but at the same time, their skills are remarkable, effective, and constructive. The more positive interpretation may feel like revisionism, but it is the daily experience of many female speakers, particularly in their conversations with other females.

4.5. Conclusion: Possible Ramifications

The difficulty with the above positive valuation of females' politeness strategies is that even when it is intended as such, it may not be interpreted as such. Stereotyping can cause assessments that are negative and even ones that contradict the acoustics of what female speakers actually produce. Expectations play a powerful role in perception. If listeners tend to react as if stereotypes are fact, it is important to understand them and where appropriate dispel them and describe the actual situation, including the nuanced nature of politeness strategies and what they mean as they are adapted by speakers according to their audience.

Furthermore, while comparable data from male speakers is needed to determine their levels of accommodation and if/how they adapt their pitches according to the gender and status of their interlocutors, the two genders may emphasize different politeness strategies, depending on the situation. If politeness strategies are different between the two interlocutors in a conversation, they may be misinterpreted and the conversation may not go smoothly. Erickson and Shultz (1982), Tannen (1984), and Wennerstrom (2001) have discussed how conversations can break down when different styles and expectations are at play. They all suggest that when conversations are out of synch and there is little rapport, not only is the conversation less pleasant, but the actual content of the speakers' turns may not be interpreted correctly. Communication can break down on every level. These kinds of problems may arise in conversations between any two speakers, regardless of their gender, who use different politeness styles.

Perhaps due to their differential use of politeness styles and different conversational experiences, men and women may interpret the same pitch patterns differently. In contrast to the perception studies where male and female listeners reacted completely similarly, Batstone and Tuomi (1981) found that while both male and female listeners identified the same two major factors when rating female voices on semantic differential scales, it was the "active" feminine characteristics (e.g., colorful, lively, interesting) that were more salient to the female listeners and the "passive" characteristics (e.g., soft, gentle, sweet) that were more salient to the males. The same acoustic data can create different percepts. Not only is it possible that males and females get different pitch patterns from female speakers due to the employment of different pitch strategies but they may also interpret the same patterns differently.

The finding of increased variability for my female subjects when they are speaking with other females due to the use of camaraderie or solidarity-based politeness may be especially important to consider. If they use this pattern/strategy more with females as a part of audience design as they know males and females have different expectations in conversations, it may work against them if a particular pitch pattern is positively evaluated by both genders. Since some of the research presented above suggests that both males and females find high variability effective in both male and female speakers, the results from experiment one suggest that women are showing their best side to other women more than they are to men. It may be the case that some variability is positively evaluated, but past a certain amount, it might be deemed emotional and evaluated negatively, perhaps particularly by male listeners. The waters are muddy and possibly dangerous. Speakers have to know when and how to adjust their strategies according to the social characteristics of their interlocutor and when not to, and this knowledge is difficult to come by, given the complex, multi-layered, ambivalent, and stereotyped nature of listeners' perceptions.

The interpretations of the pitch patterns found in the data that I have offered and have found reference to in the literature cannot be accepted without empirical validation. I have suggested that both increased pitch variability and pitch accommodation reflect a solidarity-based politeness strategy, and that high pitch can reflect a deference-based politeness strategy. However, these kinds of claims must be independently validated. In the next chapter, I suggest several possibilities for future research, including perceptual experiments that test whether or not these pitch patterns are in fact interpreted in these ways. This lack of and need for empirical validation exists at all levels, including interpretations of speakers' intentions. Claims about what speakers are trying to do with their pitch patterns, and with any aspect of their communicative efforts, cannot be accepted they are based on impressionistic evidence. Interpretations of the motivations for pitch patterns must be independently and empirically tested before we can assess the competing explanations.

Chapter 5. Conclusion

5.1. Summary of Results

This dissertation addresses the question of how speakers communicate who they are via their pitch patterns and the politeness styles they convey and how their strategies vary depending on the social characteristics of their interlocutor. The two experiments conducted found evidence of pitch patterns indicative of both positive and negative politeness strategies in the female subjects. I have interpreted the higher pitch variability between females and the accommodation found with both males and females as evidence of solidarity or camaraderie politeness and the higher pitch with professors and the pitch divergence found in a few interviews in experiment two as evidence of deference or distance politeness. The patterns consistent with positive politeness were far more numerous than those consistent with negative politeness; this skew suggests that positive face needs among these female speakers are extremely important.

My findings demonstrate that gender and status of the interlocutor do indeed affect the pitch patterns of my subjects. These interlocutor-based differences are evidence that adaptation to the addressee occurs at the level of prosody and within the same dialect or speech community, in addition to at the level of segmental phonology and between bilingual or bi-dialectal speakers as previous research has found. They highlight the importance of studying language in its different contexts as these contexts affect the language that occurs at all of its levels. This understanding of the importance of social influences at all levels of language—including prosody—needs to considered as researchers create hypotheses and experimental designs.

5.2. Future Work

My findings provide a starting point for understanding interlocutor effects on speakers' pitch. However, other questions remain unanswered and much related and additional work remains to be done.

5.2.1. Further Pitch Analysis

The study of how pitch patterns reflect politeness strategies and how the strategies used depend on the interlocutor and social situation cannot be based on data from female speakers alone. Gender and language studies are about how gender is constructed by men and women. I focused on women's pitches because it may be particularly important to study them as women have been historically understudied and have so many stereotypes about their speech at play. However, the questions of whether male speakers adjust their pitches according to the gender and status of their interlocutor and whether or not (and how) they accommodate their pitch patterns remain. It is probable that there are many patterns and strategies among male speakers, just as there are among female speakers, and it is important to find them and clarify their distribution, as well. Furthermore, it is not assumed that the same patterns or distribution of patterns will be found in other dialects (involving other regions and/or ethnicities, for example) or in other sexual orientations or other age groups. It is my working assumption that the communicative strategies that express and construct gender do so as it interacts with other social identities within individuals.

These communicative strategies depend on the precise social situation and what speakers are trying to communicate at a given instant, however. My first experiment took these moment-by-moment factors into account by manipulating the topics of conversation. Three kinds of topic were included in the interviews, but only the first two, explicitly gendered topics and mundane topics, were included in the analysis and there were no significant findings based on them. The third group of topics was designed to elicit a face-threatening act with the idea that when a female speaker commits a speech act that is aggressive in its content, she may soften it with femininity cues and/or certain politeness strategies conveyed via pitch or voice quality. This line of inquiry was not explored in this dissertation and remains for future research. To investigate it fully, pitch patterns in small portions of a conversation will need to be examined and tied to either speech acts or stances and attitudes towards the situation, interlocutor, etc. (cf. Goffman's 1981 discussion of footing which suggests speakers' alignments towards the situation can be communicated via prosody). It may be possible to determine how exactly pitch variation is tied to the speech content or to the politeness strategy being used via close discourse and pitch analysis. The experiments in my dissertation explore the global changes in pitch tessitura and variability over time that depend on the characteristics of the interlocutor; the next step is to explore the specific changes within a small segment of a conversation to determine their nature and whether or not they also vary according the interlocutor. Some research has been done in connecting prosody, including intonation, to discourse analysis (see Wennerstrom 2001 for a review of theoretical issues and examples). This type of work can be connected to politeness studies and gender studies, as they all undoubtedly interact.

For studies of all groups and for both global and more specific, discourse-based studies, the underlying question is why different speakers choose different pitch patterns and politeness strategies and why they focus on different aspects (e.g., gender or status) of their interlocutor. Politeness strategies and the ways they are communicated are situated in social groups and their social settings. Ethnographic studies, such as Eckert (2000), are needed in order to uncover the situated meanings of the strategies used by particular individuals. The array of individual motivations can be determined by examining how particular individuals use pitch to define themselves within their social groups in their particular surroundings.

5.2.2. Perceptual Studies

While speakers' intentions are interesting regardless of their communicative effects, a crucial part of the puzzle lies in discovering and describing the effects of their communicative efforts. Perhaps the most important sociophonetic ramification of a symbolic signal such as pitch is its effect on hearers. What do the pitch patterns in evidence convey? Assessing the consequences of pitch patterns requires understanding how hearers react to them. A perception study needs to be conducted in the future to try to measure what the pitch patterns found in my acoustic analysis communicate to listeners within the same speech community. To this end, I have designed perceptual experiments and will briefly describe them here.

The stimuli will be excerpts from my recorded conversations processed to distort the speech content but preserve the fundamental frequency (pitch). Possibilities for preserving pitch contours but obscuring speech content include low-pass filtering the speech excerpts at approximately the level of the first formant frequency. This method has the benefit of preserving the naturalness of the stimuli, but may not completely filter out other information. Another option involves extracting the pitch contours and then resynthesizing them artificially. This method would do a better job of isolating the perceptual effects of pitch, but the stimuli would be less natural. The use of both of these methods would provide more than one kind of information and may provide a clearer picture than either method used alone.

Listeners in the experiments should be from the same general speech community as the recorded subjects, in this case, Caucasian, heterosexual, native speakers of California English, ages 18-23. Two kinds of processed stimuli will be used. The first kind will be excerpts of speech material from one speaker at a time that are representative of the patterns found, i.e., with higher and lower median pitch and pitch variability. The naïve listener-subjects will rate them along scales of friendliness and deference to help answer the question of whether the hypothesized politeness strategies are successful or at least communicating what was hypothetically intended. The second kind of stimuli will be excerpts including speech material from both interlocutors with different levels and kinds of pitch accommodation, and the listener-subjects will rate the comfort level of the conversation, i.e., whether or not it was a "good" conversation and how well the participants were getting along. This kind of rating might help answer the question of whether conversations where interlocutors accommodate to each other are smoother or more successful conversations. Another possible kind of rating involves judging the personal attributes of the speakers, as it may be the case that users of different politeness strategies and pitch patterns or even different levels of pitch accommodation are perceived to have different traits in the realms of intelligence, friendliness, etc. Researchers have done some similar perceptual studies using resynthesized speech (cf.

Brown, Strong, and Rencher, 1974, and van Bezooijen, 1995, 1996), but they have tended to use read speech or to focus on the resynthesis of one sentence. Experiments such as those I am proposing will provide ratings of excerpts reflective of the pitch patterns found in natural conversations.

The perceptual side of the study suggests the importance of such sociophonetic studies. Our communicative strategies are perceived and have an effect that may or may not be what we intend. It is necessary to clarify the functions of pitch in order to uncover the social implications of the messages we are sending through the use of different pitch patterns.

Both perceptual and acoustic experiments need to be performed to determine whether or not or how voice quality, in particular, breathiness, is manipulated as a symbolic tool to convey gender identity, stance, etc., as well. Since voice quality may well be important in the phonetic construction of femininity in different contexts, both the literature and the issues leading to the design of an acoustic study of breathiness, as well as the difficulties such a study entails, are summarized in Appendix 7.

5.3. Conclusions

The experiments in this dissertation provide a concrete start towards understanding how speakers use pitch to convey how they view themselves and their interlocutor within a conversation. The most robust trend in the results of experiment one, which measured subjects' pitches (median pitch, standard deviation of pitch, and pitch range) in interviews with a male and female peer and professor, showed that female speakers had more varied pitches when speaking with other females. The results of experiment two, which measured the pitches of both the subjects and their interlocutors in the same interviews analyzed in experiment one, showed evidence of accommodation in several of the interviews in that the pitches of the subjects and interlocutors positively correlated. Speakers do indeed adapt their pitch patterns depending on the social characteristics of their interlocutor, sometimes even adapting them towards the pitches of their interlocutor. Furthermore, the majority of these female subjects' pitch patterns can be interpreted as reflecting their use of positive politeness strategies based on camaraderie or solidarity. These experiments also demonstrate the possibility of rigorously testing sociophonetic hypotheses.

In addition to the contribution my experiments and those described in the preceding section can make to the scholarly understanding of prosody in audience design, the acoustic study of speakers' pitch characteristics is important because hearers react to pitch, and how hearers react determines the courses of conversations, and conversations shape our lives. The ramifications of pitch patterns are multi-layered and complex. They may or may not be interpreted as intended. Not only are high pitch and high variability of pitch associated with negative qualities at times (e.g., lack of competence or emotionality), but stereotypes can color perceptions to the extent that their influence outweighs that of the pitch patterns actually present. Even when listeners react to the patterns as they occur, they make judgments about speakers based on their pitch patterns and the politeness strategies they communicate, and when politeness strategies conflict within a conversation, the conversation may not go well. As Erickson and Shultz (1982) discuss, in situations where one person has the authority to influence her or his interlocutor's fate, e.g., in a job interview, the consequences for individuals can be profound.

If linguists can shed any light on the situation or lessen the possibility of miscommunication, they can clarify what pitch patterns occur and how they might be intended in different situations. This kind of research is important because it can help to dispel stereotypes that drive perceptions of speakers despite having no basis in reality. Awareness of the variety of politeness strategies and the existence of such stereotypes may help avert unmerited negative perceptions of some female speakers. However, both stereotypes and an individual's ideas about what kind of politeness is appropriate in a given situation are extremely difficult to challenge.

In addition to shedding light on what strategies female speakers use in different situations, these two experiments also contribute to the body of knowledge on social influences on female speakers' pitch. They also demonstrate that the study of prosodic and even paralinguistic aspects of language, like the study of all aspects of language, cannot be divorced from the social settings in which they occur. By elucidating the nature of their social pitch manipulations that are based on expectations of how to appropriately speak to different interlocutors and on accommodation, and the extent of these manipulations, these experiments contribute data towards determining how society and biology both explain female speakers' use of pitch and how and why their pitches differ from males' pitches. Not only do many factors contribute but they also interact with each other. Male and female larynxes give them different kinds of voices both between and within the two genders. The two genders have a variety of politeness strategies available to them, but their interlocutors expect different behaviors from them based on their gender, and they know that different strategies are useful or appropriate depending on their interlocutor and the social situation. Individuals within each gender also make choices about the kind of person/female/male they want to be as they actively construct their gendered selves throughout all of their linguistic and other actions. The reactions to and interpretations of their communications and strategies are gendered and may or may not be based on what is actually intended. These interwoven, interacting, and sometimes conflicting layers make their elucidation extremely challenging, but each piece of knowledge that is gathered contributes a piece to the puzzle.

The puzzle will be completed with the addition of each piece of knowledge, but also because each piece might inspire other researchers to respond to it—to expose its errors, carry it farther, answer questions posed, and uncover new questions. At the very least, it is my hope that my efforts will demonstrate that the study of the interactionally-based variation of pitch and voice quality is a valid pursuit; these are manipulable symbolic cues and humans, opportunists that they are, use them, and the consequences of their communicative choices over time may indeed be profound.

References

Aron, A. and E.N. Aron. 1999. *Statistics for Psychology*. Upper Saddle River, NJ: Prentice-Hall.

Aronovitch, C.D. 1976. The voice of personality: Stereotyped judgments and their relationship to voice quality and sex of speaker. *The Journal of Social Psychology*, 99: 207-220.

Awan, S.N. and P.B. Mueller. 1992. Speaking fundamental frequency characteristics of centenarian females. *Clinical Linguistics & Phonetics*, 6(3): 249-254.

Batstone. S. and S.K. Tuomi. 1981. Perceptual characteristics of female voices. *Language and Speech*, 24(2): 111-123.

Bell, A. 1984. Language style as audience design. *Language in Society*, 13: 145-204.

Biemans, M. 1998. The effect of biological gender (sex) and social gender (gender identity) on three pitch measures. In Van Bezooijen, R. and R. Kager (Eds.), *Linguistics in the Netherlands 1998*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Biemans, M. and R. van Bezooijen. 1996. The effect of masculinity and femininity on pitch. In Warner, N., J. Ahlers, L. Bilmes, M. Oliver, S. Wertheim, and M. Chen (Eds.), *Gender and Belief Systems: Proceedings of the Fourth Berkeley Women and Language Conference*. Berkeley, CA: Berkeley Women and Language Group. 25-36.

Blankenship, B. 1997. *The Time Course of Breathiness and Laryngealization in Vowels*. PhD Dissertation, University of California, Los Angeles.

Boersma, P. 1993. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Institute of Phonetic Sciences. University of Amsterdam. Proceedings*, 17: 97-110.

Brazil, D. 1985. *The Communicative Value of Intonation*. Discourse Analysis Monograph 8. Birmingham: University of Birmingham English Language Research.

Brend, R. 1975. Male-female intonation patterns in American English. In Thorne, B. and N. Henley (Eds.), *Language and Sex: Difference and Dominance*. Rowley, MA: Newbury House Publishers. 84-87.

Brouwer, D. 1982. The influence of the addressee's sex on politeness in language use. *Linguistics*, 20: 697-711.

Brouwer, D., M. Gerritsen, and D. de Haan. 1979. Speech differences between women and men: On the wrong track? *Language in Society*, 8: 33-50.

Brown, B.L., Strong, W.J., and A.C. Rencher. 1974. Fifty-four voices from two: The effects of simultaneous manipulations of rate, mean fundamental frequency, and variance of fundamental frequency on ratings of personality from speech. *Journal of the Acoustical Society of America*, 55(2): 313-318.

Brown, P. and S.C. Levinson. 1987. *Politeness: Some Universals in Language Usage*. Cambridge: Cambridge University Press.

Cameron, D. 1985. *Feminism and Linguistic Theory*. London: Macmillan.

Chen, R. 2001. Self-politeness: A proposal. *Journal of Pragmatics*, 33: 87-106.

Chiba, T. and Kajiyama, M. 1941. *The Vowel—Its Nature and Structure*. Tokyo.

Coleman, R.O. 1971. Male and female voice quality and its relationship to vowel formant frequencies. *Journal of Speech and Hearing Research*, 14: 565-577.

Coleman, R. 1976. A comparison of the contributions of two voice quality characteristics to the perception of maleness and femaleness in the voice. *Journal of Speech and Hearing Research*, 19: 168-180.

Connor-Linton, J. 1986. Gender differences in politeness: The struggle for power among adolescents. In Connor-Linton, J., Hall, C.J., and M. McGinnis (Eds.), *Social and Cognitive Perspectives on Language*. Southern California Occasional Papers in Linguistics, Volume 11. Department of Linguistics, University of Southern California.

Coupland, N. 1985. 'Hark, hark, the lark': Social motivations for phonological style-shifting. *Language and Communication*, 5(3): 153-171.

de Pinto, O. and H. Hollien. 1982. Speaking fundamental frequency characteristics of Australian women: Then and now. *Journal of Phonetics*, 10: 367-375.

Durkheim, E. 1915. *The Elementary Forms of the Religious Life*. London.

Eckert, P. 2000. *Linguistic Variation as Social Practice*. Oxford: Blackwell.

Eckert, P. and S. McConnell-Ginet. 1992. Think practically and look locally: Language and gender as community-based practice. *Annual Review of Anthropology*, 21: 461-490.

Edelsky, C. 1979. Question intonation and sex roles. *Language in Society*, 8: 15-32.

Eklund, I. and H. Traunmüller. 1997. Comparative study of male and female whispered and phonated versions of the long vowels of Swedish. *Phonetica*, 54 (1): 1-21.

Epstein, M.A. 2001. *Voice Quality and Prosody in English*. UCLA Dissertation Prospectus.

Erickson, F. and J. Shultz. 1982. *The Counselor as Gatekeeper*. NY: Academic Press.

Fant, G. 1966. A note on vocal tract size factors and non-uniform f-pattern scalings. *Quarterly Progress and Status Report*, 4. Stockholm, Royal Institute of Technology, Dept. of Speech Communication, Speech Transmission Laboratory. 22-35.

Fay, P.J. and W.C. Middleton. 1940. Judgment of Kretschmerian body types from the voice as transmitted over a public address system. *The Journal of Social Psychology*, 12: 151-162.

Finegan, E. and D. Biber. 1994. Register and social dialect variation: An integrated approach. In Biber, D. and E. Finegan (Eds.), *Sociolinguistic Perspectives on Register*. NY, NY: Oxford University Press. 315-347.

Forchhammer, E. 1974. *Stemmens funktioner og fejlfunktioner*. Copenhagen: Munsgaard.

Giles, H. and P.F. Powesland. 1975. *Speech Style and Social Stratification*. NY, NY: Academic Press Inc.

Goffman, E. 1967. *Interaction Ritual: Essays on Face to Face Behavior*. Garden City, NY.

Goffman, E. 1981. *Forms of Talk*. Philadelphia: University of Pennsylvania Press.

Goldstein, U. 1980. *An Articulatory Model for the Vocal Tracts of Growing Children*. PhD Dissertation, MIT. Cambridge, MA.

Graddol, D. 1986. Discourse specific pitch behavior. Chapter 11 in Johns-Lewis, C. (Ed), *Intonation in Discourse*. San Diego, CA: College-Hill Press, Inc.

Graddol, D. and J. Swann. 1983. Speaking fundamental frequency: Some physical and social correlates. *Language and Speech*, 26(4): 351-366.

Graddol, D. and J. Swann. 1989. *Gender Voices*. Cambridge, MA: Basil Blackwell.

Greif, E. 1980. Sex differences in parent-child conversations. In: Kramarae, C. (Ed.), *The Voices and Words of Women and Men*. NY: Pergamon Press.
Günzburger, D. 1996. Gender adaptation in the speech of transsexuals: From sex transition to gender transmission? In Warner, N., J. Ahlers, L. Bilmes, M. Oliver, S. Wertheim, and M. Chen, Eds., *Gender and Belief Systems: Proceedings of the Fourth Berkeley Women and Language Conference*. Berkeley, CA: Berkeley Women and Language Group. 269-278.
Haan, J. and V.J. van Heuven. 1999. Male vs. female pitch range in Dutch questions. *Proceedings of the 14th International Congress of Phonetic Sciences. San Francisco*. 1581-1584.
Hanson, H.M. 1997. Glottal characteristics of female speakers: Acoustic correlates. *JASA*, 101(1): 466-481.
Hanson, H.M. and E.S. Chuang. 1999. Glottal characteristics of male speakers: Acoustic correlates and comparison with female data. *JASA*, 106(2): 1064-1077.
Henley, N.M., K. Meng, D. O'Brien, W.J. McCarthy, and R.J. Sockloskie. 1998. Developing a scale to measure the diversity of feminist attitudes. *Psychology of Women Quarterly*, 22: 317-348.
Henley, N.M., L.R. Spalding, and A. Kosta. 2000. Development of the short form of the Feminist Perspectives Scale. *Psychology of Women Quarterly*, 24: 254-256.
Henton, C.G. 1989. Fact and fiction in the description of female and male pitch. *Language and Communication*, 9 (4): 299-311.
Henton, C.G. 1995. Pitch dynamism in female and male speech. *Language and Communication*, 15 (1): 43-61.
Henton, C.G. 1999. Where is female synthetic speech? *Journal of the International Phonetic Association*, 29(1): 51-61.
Henton, C.G. and A.W. Bladon. 1985. Breathiness in normal female speech: Inefficiency versus desirability. *Language and Communication*, 5(3): 221-227.
Henton, C.G. and A.W. Bladon. 1988. Creak as a sociophonetic marker. In Hyman, L.M., and C.N. Li (Eds.), *Language, Speech and Mind: Studies in Honor of Victoria A. Fromkin*. London: Routledge. 3-29.
Hillenbrand, J., Cleveland, R.A., and R.L. Erickson. 1994. Acoustic correlates of breathy vocal quality. *Journal of Speech and Hearing Research*, 37: 769-778.
Hillenbrand, J. and R.A. Houde. 1996. Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech. *Journal of Speech and Hearing Research*, 39: 311-321.
Hollien, H. and B. Jackson. 1973. Normative data on the speaking fundamental frequency characteristics of young adult males. *Journal of Phonetics*, 1: 117-120.
Holmberg, E.B., Hillman, R.E., Perkell, J.S., Guiod, P.C., and S.L. Goldman. 1995. Comparisons among aerodynamic, electroglottographic, and acoustic spectral measurements of female voice. *Journal of Speech and Hearing Research*, 38: 1212-1223.
Holmes, J. 1993. New Zealand women are good to talk to: An analysis of politeness strategies in interaction. *Journal of Pragmatics*, 20: 91-116.
Jacobi, J. 1996. *The Vocal Advantage*. Englewood Cliffs, NJ: Prentice Hall.
Jary, M. 1998. Relevance theory and the communication of politeness. *Journal of Pragmatics*, 30: 1-19.

Johnstone, B., Ferrara, K., and J. Mattson Bean. 1992. Gender, politeness, and discourse management in same-sex and cross-sex opinion-poll interviews. *Journal of Pragmatics*, 18: 405-430.
Jassem, W. 1971. Pitch and compass of the speaking voice. *JIPA*, 1: 59-68.
Kahane, J. 1978. A morphological study of the human prepubertal and pubertal larynx. *American Journal of Anatomy*, 151: 11-20.
Karlsson, I. 1986. Glottal wave forms for normal female speakers. *Journal of Phonetics*, 14 (3/4): 415-419.
Karlsson, I. and M. Rothenberg. 1992. Intercultural variations in gender-based language differences in young children. *Quarterly Progress and Status Report*, 1. Stockholm, Royal Institute of Technology, Dept. of Speech Communication, Speech Transmission Laboratory. 1-17.
Kasper, G. 1990. Linguistic politeness: Current research issues. *Journal of Pragmatics*, 14: 193-218.
Klatt, D.H. and L.C. Klatt. 1990. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *JASA*, 87(2): 820-857.
Kreiman, J. and B. Gerratt. 2000. Measuring vocal quality. Chapter 7 in Kent, R.D. and M.J. Ball (Eds.), *Voice Quality Measurement*. San Diego, CA: Singular Publishing Group.
Künzel, H.J. 1989. How well does average fundamental frequency correlate with speaker height and weight? *Phonetica*, 46(1-3): 117-125.
Ladefoged, P. 1982. *A Course in Phonetics*. NY, NY: Harcourt Brace Jovanovich, Inc.
Ladefoged, P. and N. Antoñanzas-Barroso. 1985. Computer measures of breathy voice quality. *UCLA Working Papers in Phonetics*, 61: 79-86.
Lakoff, R.T. 1973. The logic of politeness; or, minding your p's and q's. *Proceedings of the Ninth Regional Meeting of the Chicago Linguistics Society*, 292-305.
Lakoff, R.T. 1975. *Language and Woman's Place*. NY, NY; Harper & Row.
Lakoff, R.T. 1990. *Talking Power*. Basic Books.
Lass, N.J., Kelley, D.T., Cunningham, C.M., and K.T. Sheridan. 1980. A comparative study of speaker height and weight identification from voiced and whispered speech. *Journal of Phonetics*, 8: 195-204.
Laver, J. 1991. *The Gift of Speech. Papers in the Analysis of Speech and Voice.* Edinburgh: Edinburgh University Press.
Laver, J. and P. Trudgill. 1979. Phonetic and linguistic markers in speech. In Scherer, K.R. and H. Giles (Eds.), *Social Markers in Speech*. Cambridge: Cambridge University Press.
Lieberman, P. 1967. *Intonation, Perception, and Language*. Research Monograph No. 28. Cambridge, MA: The MIT press.
Linke, C.E. 1973. A study of pitch characteristics of female voices and their relationship to vocal effectiveness. *Folia Phoniatrica*, 25: 173-185.
Loveday, L. 1981. Pitch, politeness and sexual role: An exploratory investigation into the pitch correlates of English and Japanese politeness formulae. *Language and Speech*, 24(1): 72-89.
Majewski, W., Hollien, H., and J. Zalewski. 1972. Speaking fundamental frequency of Polish adult males. *Phonetica*, 25: 119-125.

Markel, N.N., Prebor, L.D., and J.F. Brandt. 1972. Biosocial factors in dyadic communication: Sex and speaking intensity. *Journal of Personality and Social Psychology*, 23(1): 11-13.
Mattingly, I. 1966. Speaker variation and vocal-tract size. *JASA*, 39(6): 1219. (Abstract.)
McConnell-Ginet, S. 1978. Intonation in the social context: Language and sex. Paper presented at the 9th International Congress on Sociology, Uppsala, Sweden.
McConnell-Ginet, S. 1983. Intonation in a man's world. In Thorne, B., Kramarae, C., and N. Henley (Eds.), *Language, Gender, and Society*. Rowley, MA: Newbury House. 69-88.
Meier, A.J. 1995. Passages of politeness. *Journal of Pragmatics*, 24: 381-392.
Moore, B.C.J. 1999. Aspects of auditory processing related to speech perception. Chapter 17 in Hardcastle, W. and J. Laver (Eds.), *The Handbook of Phonetic Sciences*. Malden, MA: Blackwell Publishers.
Moore, K. 1995. Identification of male and female voice qualities in pre-pubescent children. *Proceedings of the 13th International Congress of Phonetic Sciences*. Stockholm. Vol. 2. 298-301.
Morse, J. 2002. Reporters' notebook. *Time* magazine, April 1, 2002 issue.
Morton, E.S. 1977. On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *The American Naturalist*, 111(981): 855-869.
Nairn, M.J. 1995. The perception of gender differences in the speech of 4 1/2-5 1/2 year old children. *Proceedings of the 13th International Congress of Phonetic Sciences*. Stockholm. Vol. 2. 302-305.
Nolan, F.J. 1983. *The Phonetic Bases of Speaker Recognition*. Cambridge: Cambridge University Press.
O'Grady, K.E., J.S. Freda, and P.J. Mikulka. 1979. A comparison of the Adjective Check list, Bem Sex Role Inventory, and Personal Attributes Questionnaire masculinity and femininity subscales. *Multivariate Behavioral Research*, 14(2): 215-225.
Ochs, E. 1992. Indexing gender. In Duranti, A. and C. Goodwin (Eds.), *Rethinking Context: Language as an Interactive Phenomenon*. Cambridge: Cambridge University Press.
Ohala, J.J. 1994. The frequency code underlies the sound-symbolic use of voice pitch. In Hinton, L., Nichols, J., and J.J. Ohala (Eds.), *Sound Symbolism*. Cambridge; NY: Cambridge University Press. 325-347.
Ohala, J.J. 1996. Ethological theory and the expression of emotion in the voice. *Proceedings of the 4th International Conference on Spoken Language Processing* (Philadelphia). Wilmington: University of Delaware. Vol. 3. 1812-1815.
Ohara, Y. 1992. Gender-dependent pitch levels: A comparative study in Japanese and English. In Hall, K., Bucholtz, M. and B. Moonwomon, Eds., *Locating Power. Proceedings of the Second Berkeley Women and Language Conference*. Berkeley, CA: Berkeley Women and Language Group. Volume 2, 468-477.
Ohara, Y. 1999. Performing gender through voice pitch: A cross-cultural analysis of Japanese and American English. In Pasero, U., and F. Braun (Eds.), *Perceiving and Performing Gender [Wahrnehmung und Herstellung von Geschlecht]*. Opladen/Wiesbaden: Westdeutscher Verlag. 105-116.
Peterson, G.E. and H.L. Barney. 1952. Control methods used in a study of the vowels. *JASA*, 24: 175-184.

Pike, K.L. 1945. *The Intonation of American English.* Ann Arbor: University of Michigan Press.
Price, P.J. 1989. Male and female voice source characteristics: Inverse filtering results. *Speech Communication*, 8: 261-277.
Rickford, J.R. and F. McNair-Knox. 1994. Addressee- and topic-influenced style shift: A quantitative sociolinguistic study. In Biber, D. and E. Finegan (Eds.), *Sociolinguistic Perspectives on Register.* NY, NY: Oxford University Press. 235-276.
Sachs, J., Lieberman, P. and D. Erickson. 1973. Anatomical and cultural determinants of male and female speech. In Shuy, R.W. and R.W. Fasold (Eds.), *Language Attitudes: Current Trends and Prospects.* Washington, D.C.: Georgetown University Press. 74-84.
Sachs, J. 1975. Cues to the identification of sex in children's speech. In Thorne, B. and N. Henley (Eds.), *Language and Sex: Difference and Dominance.* Rowley, MA: Newbury House Publishers. 152-171.
Schiffrin, D. 1994. *Approaches to Discourse.* Malden, MA: Blackwell Publishers Inc.
Sheldon, A. 1992. Conflict talk: Sociolinguistic challenges to self-assertion and how young girls meet them. *Merrill-Palmer Quarterly*, 38(1): 95-117.
Shevchenko, T.I. 1999. The sociocultural value of F0-variation in British and American English. *Proceedings of the 14th International Congress of Phonetic Sciences.* San Francisco. 1609-1612.
Smith, C.L. 1999. Marking the boundary: Utterance-final prosody in French questions and statements. *Proceedings of the 14th International Congress of Phonetic Sciences.* San Francisco. 1181-1184.
Smith, G.A. and A.J. Robinson. 2000. Segmentation of a speech waveform according to glottal open and closed phase using an autoregressive-HMM. *Proceedings of the ICSLP 2000.*
Smith, G.A. and A.J. Robinson. 2001. Segmentation of speech waveforms according to open and closed phases using duration modelling. *Proceedings of WISP 2001.*
Smith, P.M. 1985. *Language, the Sexes and Society.* New York: Basil Blackwell Inc.
Smith, S.B. 1978. *The Relationship between the Angle of the Thyroid Laminae and Vocal Fold Length.* Masters Thesis, University of Illinois. Champaign, Illinois.
Snidecor, J.C. 1940. *Studies in the pitch and duration characteristics of superior speakers.* PhD dissertation. Iowa City.
Södersten, M. and P.-å. Lindestad. 1990. Glottal closure and perceived breathiness during phonation in normally speaking subjects. *Journal of Speech and Hearing Research*, 33: 601-611.
Spence, J.T. and R.L. Helmreich. 1978. *Masculinity and Femininity: Their Psychological Dimensions, Correlates, and Antecedents.* Austin: University of Texas Press.
Stoicheff, M.L. 1981. Speaking fundamental characteristics of nonsmoking female adults. *Journal of Speech and Hearing Research*, 24: 437-441.
Stuart-Smith, J. 1999. Glasgow: Accent and voice quality. In Foulkes, P. and G. Docherty (Eds.), *Urban Voices: Variation and Change in British Accents.* London: Arnold. 203-222.

Swingle, P.G. and G.F. Hope. 1987. Interpersonal influence on temporal measures of vocalization in interview situations. *Language and Speech*, 30(2): 159-167.
Tannen, D. 1984. *Conversational Style: Analyzing Talk among Friends*. Norwood, NJ: Ablex Publishing Corporation.
Thorne, B. 1993. *Gender Play*. New Brunswick, New Jersey: Rutgers University Press.
Tielen, M.T.J. 1992. *Male and Female Speech: An Experimental Study of Sex-Related Voice and Pronunciation Characteristics*. PhD Thesis, University of Amsterdam, CIP-Gegevens Koninklijke Bibliotheek, Den Haag.
Titze, I. 1989. Physiologic and acoustic differences between male and female voices. *JASA*, 85(4); 1699-1707.
van Bezooijen, R. 1995. Sociocultural aspects of pitch differences between Japanese and Dutch women. *Language and Speech*, 38(3), 253-265.
van Bezooijen, R. 1996. The effect of pitch on the attribution of gender-related personality traits. In Warner, N., Ahlers, J., Bilmes, L., Oliver, M. Wertheim, S., and M. Chen (Eds.), *Gender and Belief Systems: Proceedings of the Fourth Berkeley Women and Language Conference*. Berkeley, CA: Berkeley Women and Language Group. 755-766.
Wennerstrom, A. 2001. *The Music of Everyday Speech: Prosody and Discourse Analysis*. Oxford, NY: Oxford University Press.
West, C. and D. Zimmerman. 1991. Doing gender. In Lorber, J. and S.A. Farrell (Eds.), *The Social Construction of Gender*. Newbury Park, CA: Sage Publications. 13-37.
Yuasa, I. 1998. Politeness strategies employed by Japanese men and women. *The Proceedings of the Fourth Annual Crossroads for Language, Interaction, and Culture*. Los Angeles: University of California, Los Angeles.
Yuasa, I. 1999. The effect of an interlocutor's gender on pitch-ranges employed by Japanese men. *Proceedings of the 14th International Congress of Phonetic Sciences*. San Francisco. 1605-1608.
Yuasa, I. 2001. Perception of pitch ranges used by American and Japanese men and women. In Benor, S., M. Rose, D. Sharma, J. Sweetland, and Q. Zhang (Eds.), *Gendered Practices in Language*. Stanford, CA: CSLI Publications.
Zemlin, W.R. 1988. *Speech and Hearing Science. Anatomy and Physiology*. Englewood Cliffs, NJ: Prentice Hall.

Appendix 1: Praat Settings Used in Experiments One and Two

Default settings in Praat's autocorrelation method (very accurate box checked) were used, except for the following parameters:

	Minimum Pitch	Octave Cost	Octave Jump Cost
Subject 1	135	0.06	0.8
Subject 2	100	0.06	0.8
Subject 3	133	0.05	0.5
Subject 4	125	0.05	0.5
Subject 5	100	0.06	0.8
Subject 6	125	0.04	0.8
Subject 7	100	0.05	0.8
Subject 8	125	0.06	0.8
Subject 9	125	0.05	0.8
Subject 10	125	0.06	0.8
Subject 11	125	0.06	0.8
Subject 12	100	0.06	0.8
Interlocutors:			
Female Professor	100	0.05	0.8
Female Peer	125	0.05	0.8
Male Professor	30	0.04	0.8
Male Peer	50	0.06	0.5

(Males' Maximum Pitch was 300.)

Appendix 2: Topic Rotation among Subjects

(See text in Chapter 2 for discussion of content of topic categories.)

Interlocutor:	female prof.	male prof.	f. peer	m. peer
Subject:				
1	A1,B1,C1	A2,B2,C2	A3,B3,C3	A4,B4,C4
2	A2,B2,C2	A1,B1,C1	A4,B4,C4	A3,B3,C3
3	A1,B1,C1	A2,B2,C2	A4,B4,C4	A3,B3,C3
4	A2,B2,C2	A1,B1,C1	A3,B3,C3	A4,B4,C4
5	A1,B1,C1	A2,B2,C2	A3,B3,C3	A4,B4,C4
6	A2,B2,C2	A1,B1,C1	A4,B4,C4	A3,B3,C3
7	A1,B1,C1	A2,B2,C2	A4,B4,C4	A3,B3,C3
8	A2,B2,C2	A1,B1,C1	A3,B3,C3	A4,B4,C4
9	A1,B1,C1	A2,B2,C2	A3,B3,C3	A4,B4,C4
10	A2,B2,C2	A1,B1,C1	A4,B4,C4	A3,B3,C3
11	A1,B1,C1	A2,B2,C2	A4,B4,C4	A3,B3,C3
12	A2,B2,C2	A1,B1,C1	A3,B3,C3	A4,B4,C4
13	A1,B1,C1	A2,B2,C2	A3,B3,C3	A4,B4,C4
14	A2,B2,C2	A1,B1,C1	A4,B4,C4	A3,B3,C3
15	A1,B1,C1	A2,B2,C2	A4,B4,C4	A3,B3,C3
16	A2,B2,C2	A1,B1,C1	A3,B3,C3	A4,B4,C4

Subjects 1-12 were female. Subjects 13-16 were male.

Appendix 3: Normalized Scores for Entire Interviews

Median

A	Sc-Med/Med
AS3	-0.040355424
AS7	-0.037511148
AS6	-0.011203008
AS10	-0.006661215
AS11	-4.36969E-05
AS8	0.013651695
AS5	0.015901268
AS12	0.030684158
AS9	0.063332388
B	
BS10	-0.034261748
BS11	-0.031618129
BS5	-0.027927412
BS4	-0.020218308
BS8	-0.012981461
BS9	-0.001422031
BS7	-0.000156545
BS12	0.00199298
BS6	0.01300722
BS3	0.034779845
J	
JS7	-0.028547805
JS6	-0.010895957
JS10	0.002797328
JS3	0.0057349
JS1	0.00635087
JS4	0.006548608
JS2	0.02128045
JS8	0.026642712
JS9	0.027776951
JS11	0.032248646
JS5	0.035255555
JS12	0.045639631
P	
PS4	-0.032735052
PS5	-0.024902041
PS2	-0.01844626
PS7	0.006738861
PS12	0.014869457

S1	Sc-Med/Med
S1J	-0.028510525
S1P	0.073361077
S2	
S2P	-0.00296648
S2J	0.058926348
S3	
S3A	-0.013772027
S3B	-0.005844788
S3J	-0.00017827
S3P	0.043274504
S4	
S4P	0.013106518
S4B	0.020589873
S4J	0.04022396
S5	
S5J	0.06907468
S5P	0.078274858
S5A	0.080998889
S5B	0.11099446
S6	
S6A	-0.015374892
S6B	-0.010588074
S6P	0.025639994
S6J	0.072736385
S7	
S7A	-0.036818254
S7B	-0.034415548
S7J	-0.006662813
S7P	0.044832557
S8	
S8B	0.025299077
S8P	0.027591909
S8A	0.043263558
S8J	0.132409919

PS8	0.016577433
PS3	0.021225616
PS10	0.023725601
PS9	0.025702362
PS1	0.034563021
PS6	0.037447227
PS11	0.04214958

S9	
S9P	-0.011203756
S9J	0.006987585
S9B	0.018685731
S9A	0.02106369
S10	
S10J	-0.004975038
S10B	0.000794001
S10A	0.015419574
S10P	0.024539898
S11	
S11J	0.000695043
S11B	0.001061638
S11A	0.004693431
S11P	0.059452034
S12	
S12J	-0.030544752
S12B	0.017479577
S12A	0.046675048
S12P	0.076518451

Standard Deviation

A	Sc-Med/Med
AS11	0.431168656
AS10	0.436977402
AS3	0.517472451
AS7	0.553967176
AS6	0.559991445
AS8	0.704652402
AS5	0.711819031
AS12	0.782192686
AS9	0.849341276
B	
BS8	0.449645905
BS9	0.453517794
BS12	0.552429383
BS11	0.558906162
BS6	0.59411523
BS3	0.599848034
BS10	0.768524026
BS5	0.891112087
BS7	1.07501672

S1	Sc-Med/Med
S1J	0.527792834
S1P	0.663269948
S2	
S2P	0.786957776
S2J	1.142723175
S3	
S3J	0.667873453
S3P	0.686088925
S3A	0.845833437
S3B	0.939822703
S4	
S4J	1.325635201
S4P	1.610248715
S4B	1.66895631
S5	
S5J	2.029524174

BS4	1.23561324

J

JS8	0.335764207
JS6	0.394404892
JS9	0.456617799
JS11	0.459771532
JS4	0.511897453
JS7	0.517589539
JS2	0.556620047
JS12	0.55922292
JS10	0.55981351
JS3	0.571658798
JS5	0.700938178
JS1	0.704415102

P

PS12	0.288532041
PS8	0.29047883
PS2	0.348644664
PS10	0.358147394
PS7	0.36842619
PS5	0.426159428
PS3	0.469981238
PS11	0.5707455
PS9	0.572981256
PS4	0.599517583
PS6	0.603487796
PS1	0.745473437

S5P	2.107244286
S5B	2.575098095
S5A	3.1592075

S6

S6B	0.400864109
S6A	0.550988625
S6J	0.55988932
S6P	0.874075184

S7

S7B	0.485800141
S7P	0.557839445
S7A	0.662069858
S7J	0.665122054

S8

S8P	0.837834533
S8B	0.887943662
S8J	1.022312797
S8A	1.026761846

S9

S9B	0.394308958
S9A	0.450372956
S9J	0.505648513
S9P	0.715350218

S10

S10J	0.779159944
S10B	0.838748298
S10A	0.854557241
S10P	0.889518881

S11

S11B	0.826130972
S11J	0.861152892
S11A	0.996633482
S11P	1.001266509

S12

S12J	0.716438122
S12B	1.252492742
S12A	1.319143676
S12P	1.967492062

Range

A	Sc-Med/Med
AS11	0.309715568
AS10	0.347169491
AS6	0.350690542
AS7	0.364705806
AS3	0.402445694
AS5	0.595765475
AS8	0.617103764
AS9	0.743193579
AS12	0.811060176
B	
BS8	0.691177578
BS6	0.712052707
BS9	0.75406769
BS11	0.765131301
BS3	0.774713612
BS12	0.871540876
BS10	1.022115214
BS5	1.12804903
BS4	1.380849107
BS7	1.494759556
J	
JS9	0.321976736
JS8	0.327373255
JS4	0.401573025
JS6	0.426552103
JS11	0.453926647
JS10	0.512939233
JS5	0.545774753
JS2	0.578326474
JS3	0.629844603
JS7	0.659548948
JS12	0.660024439
JS1	0.790357702
P	
PS12	0.251470346
PS10	0.274487426
PS8	0.328363999
PS2	0.338849776
PS5	0.348983095
PS7	0.381770112

S1	Sc-Med/Med
S1J	0.419109385
S1P	0.600052857
S2	
S2P	0.374341474
S2J	0.944655899
S3	
S3A	0.704436881
S3J	0.713919441
S3P	0.740818466
S3B	0.922203942
S4	
S4J	0.966421451
S4P	1.222243262
S4B	1.350136572
S5	
S5J	1.768500992
S5P	1.923917823
S5B	2.446684176
S5A	2.876140886
S6	
S6A	0.294808121
S6B	0.316344224
S6J	0.662140558
S6P	0.688916524
S7	
S7B	0.328697496
S7J	0.332635337
S7A	0.515968675
S7P	0.595767477
S8	
S8B	0.659631109
S8P	0.804249505
S8A	0.969105972
S8J	1.085808345

PS4	0.413595488
PS3	0.452201634
PS11	0.538098359
PS1	0.605119066
PS6	0.609523681
PS9	0.653701781

S9

S9A	0.549198562
S9J	0.556137628
S9B	0.572812433
S9P	0.743515001

S10

S10J	0.501568381
S10P	0.709048655
S10B	0.712585615
S10A	0.76195718

S11

S11B	0.508316639
S11A	0.533809669
S11J	0.581475355
S11P	0.719498136

S12

S12J	0.443137458
S12B	0.899946994
S12A	0.950139182
S12P	1.675138803

Appendix 4: Gender and Status Group Correlation Scatterplots

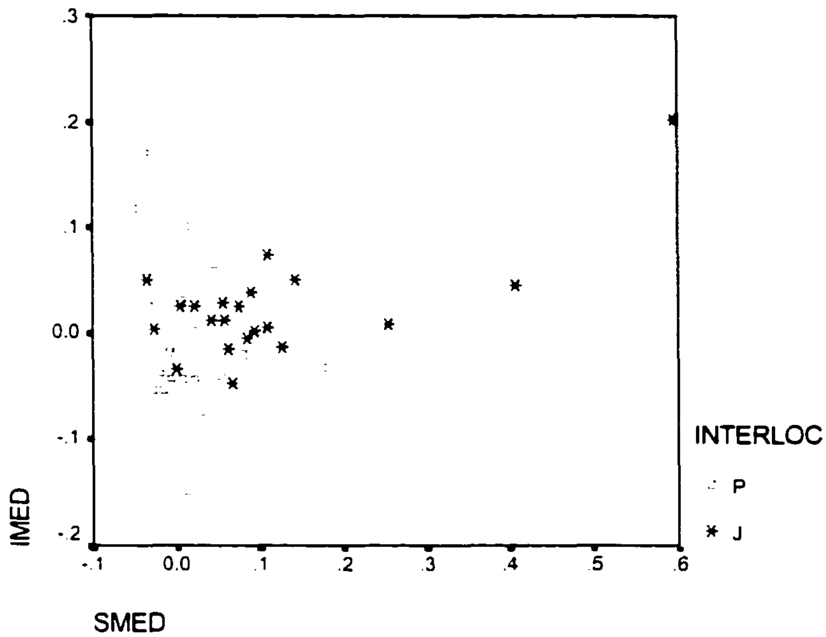


Figure 1. S2 with professors, normalized scores, medians.

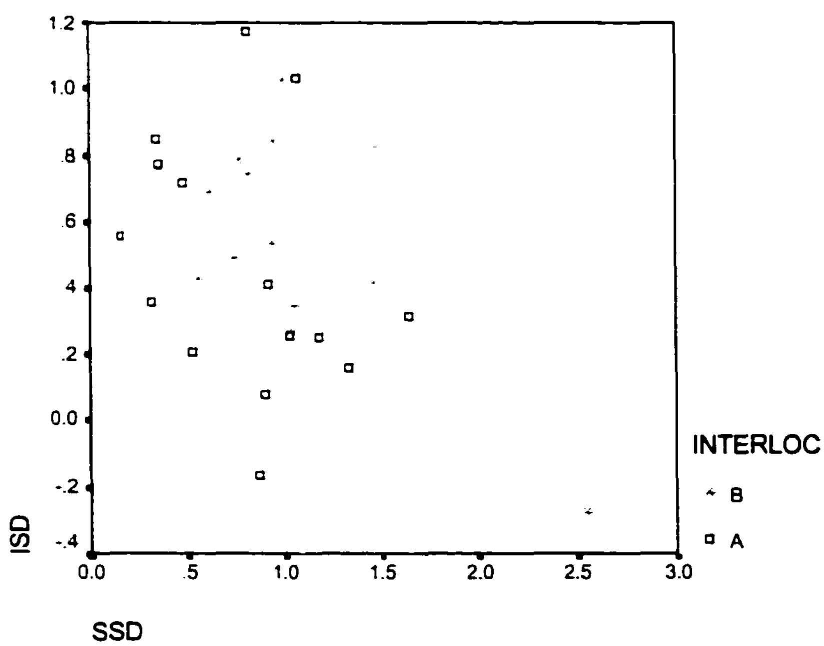


Figure 2. S3 with peers, normalized scores, SDs.

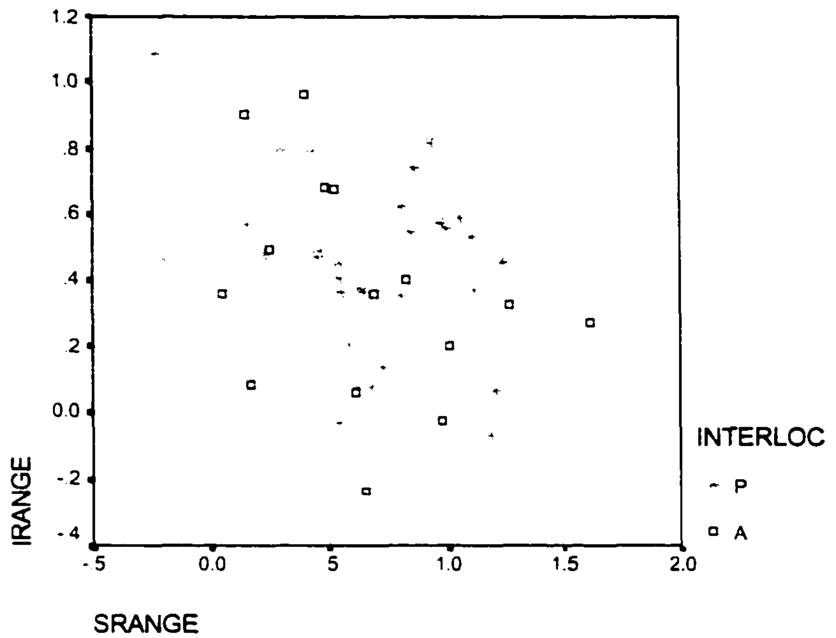


Figure 3. S3 with females, normalized scores, ranges.

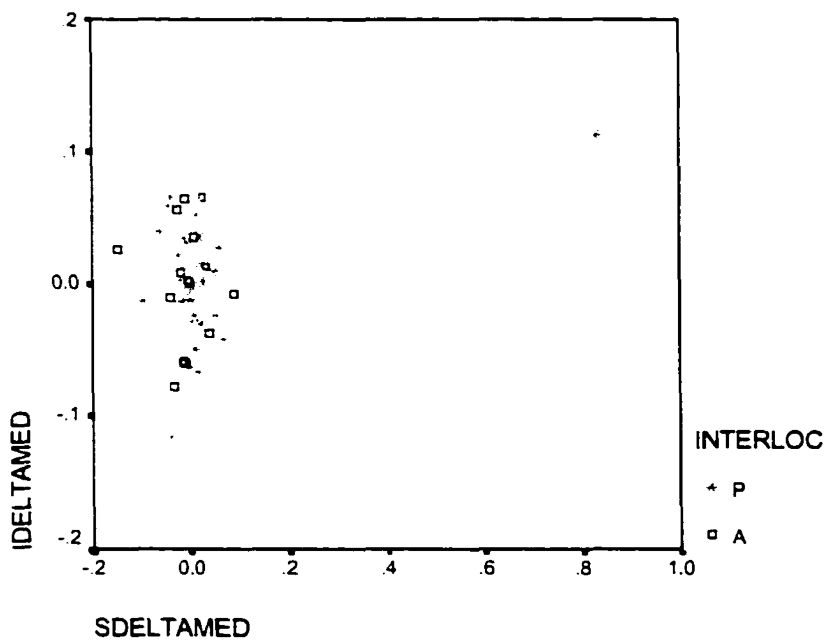


Figure 4. S3 with females, deltas, medians.

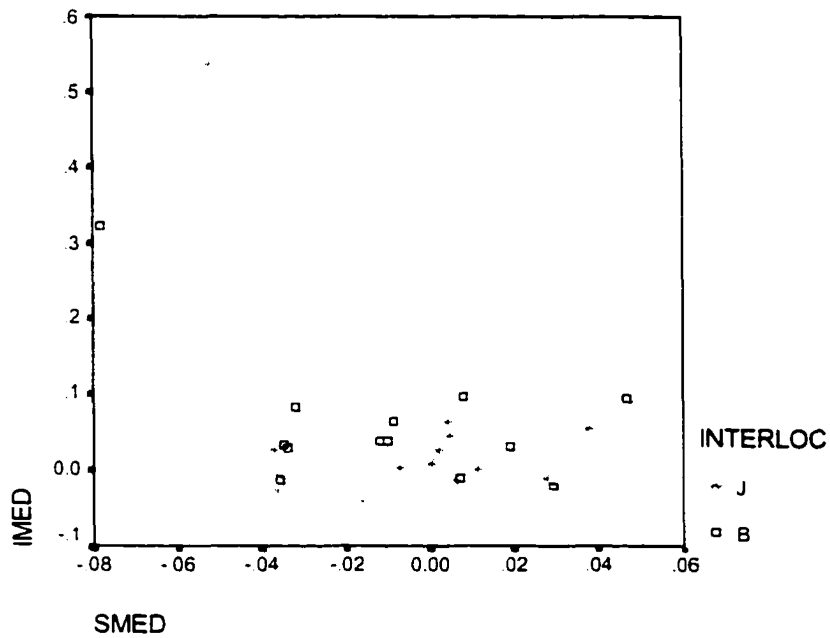


Figure 5. S3 with males, normalized scores, medians.

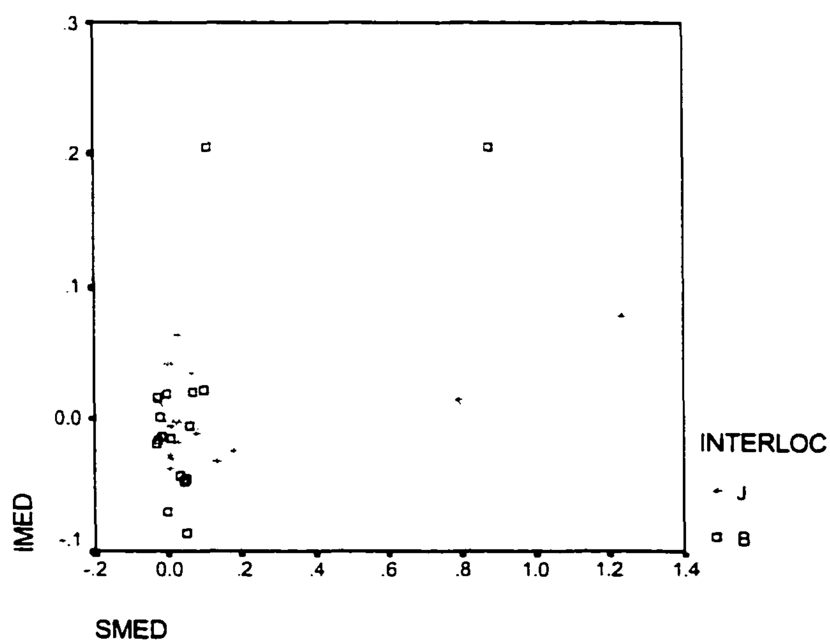


Figure 6. S4 with males, normalized scores, medians.

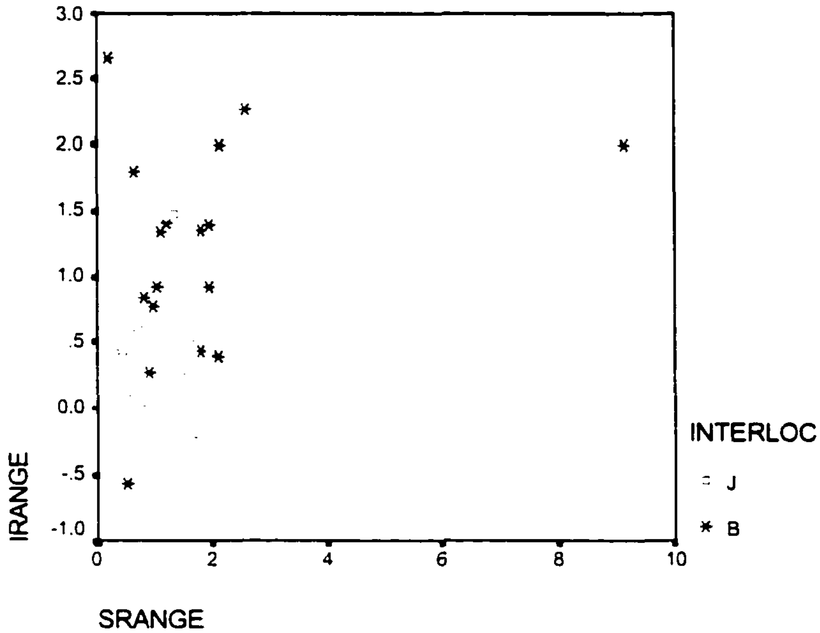


Figure 7. S4 with males, normalized scores, ranges.

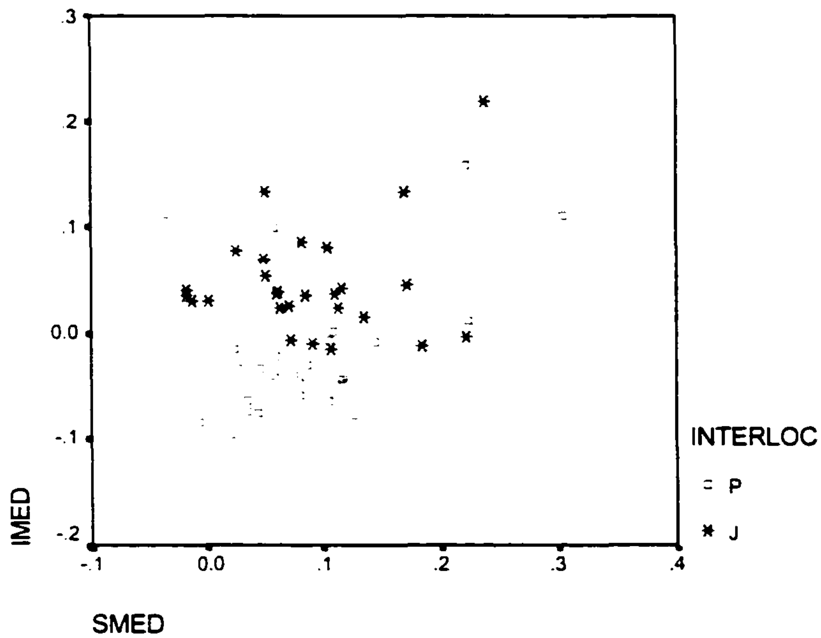


Figure 8. S5 with professors, normalized scores, medians.

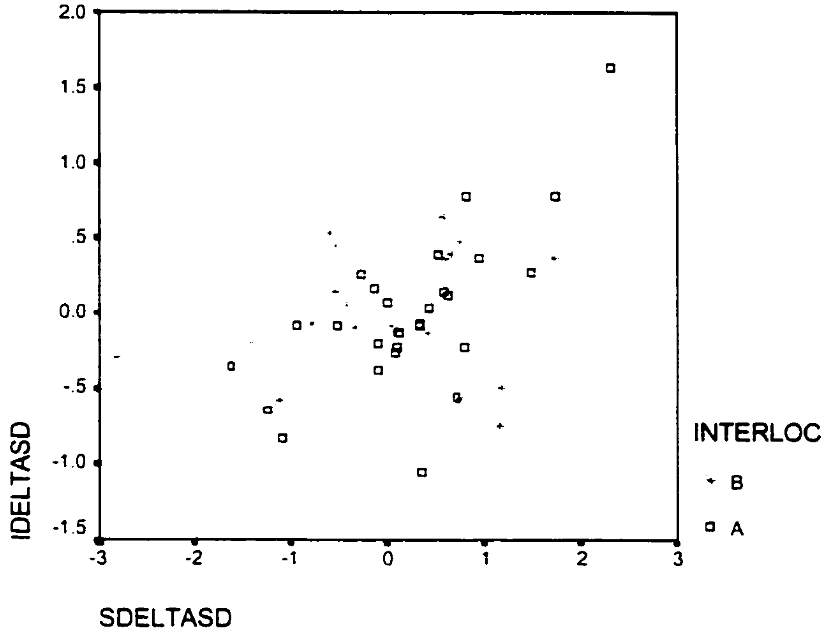


Figure 9. S5 with peers, deltas, SDs.

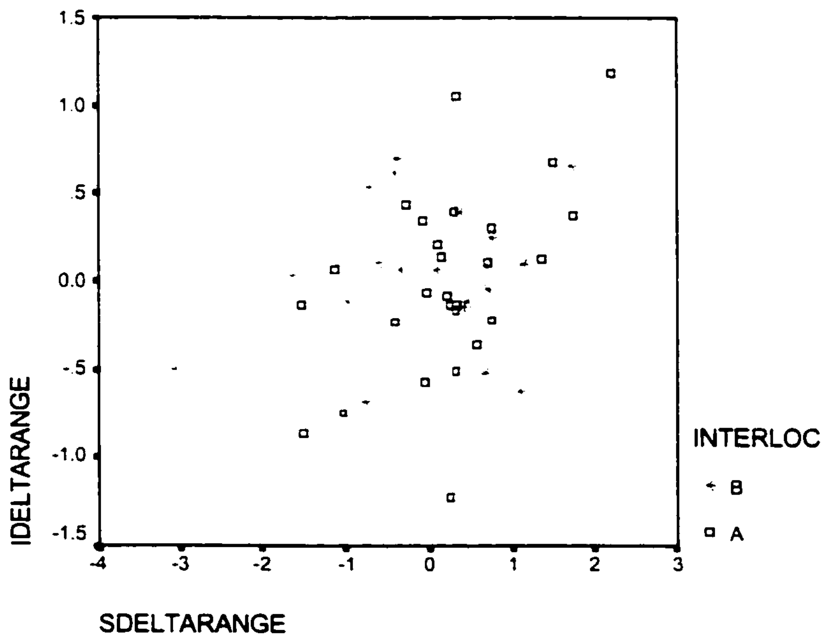


Figure 10. S5 with peers, deltas, ranges.

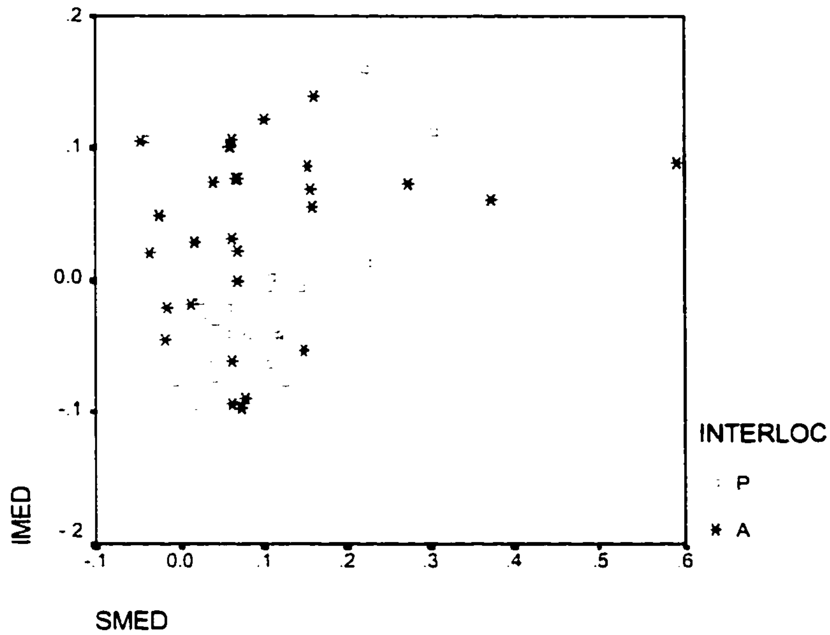


Figure 11. S5 with females, normalized scores, medians.

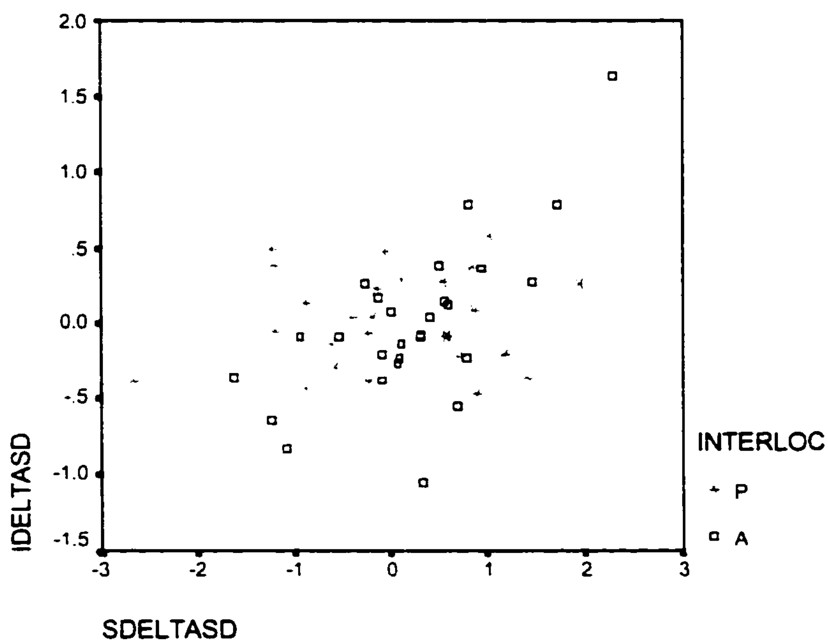


Figure 12. S5 with females, deltas, SDs.

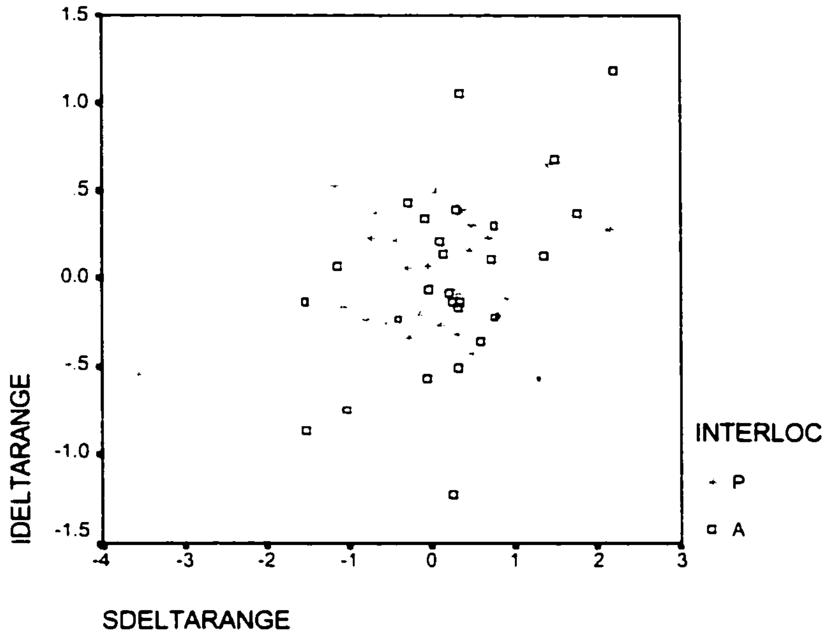


Figure 13. S5 with females, deltas, ranges.

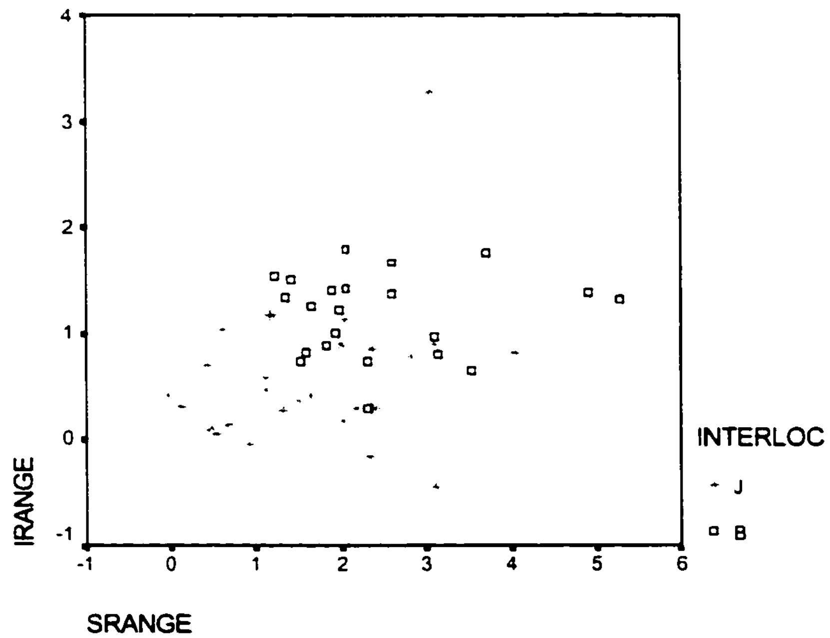


Figure 14. S5 with males, normalized scores, ranges.

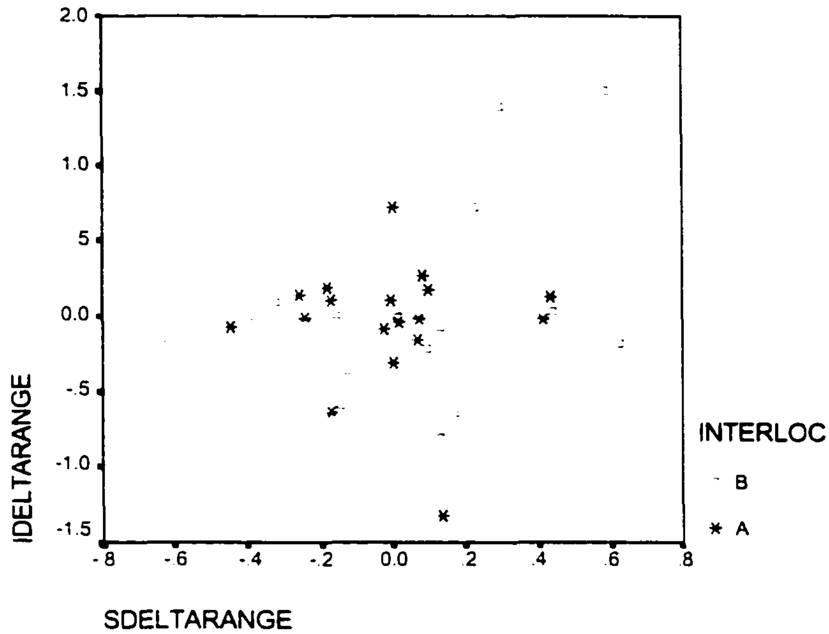


Figure 15. S6 with peers, deltas, ranges.

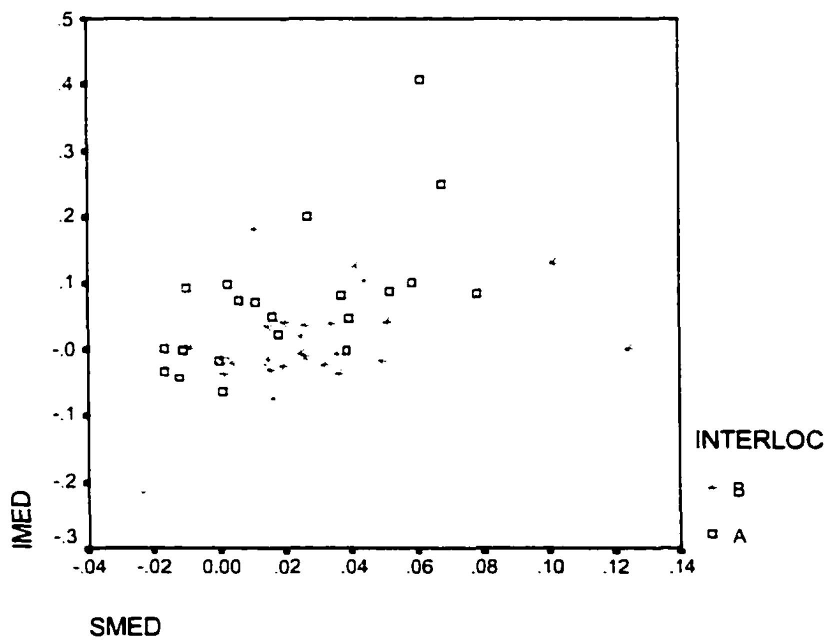


Figure 16. S9 with peers, normalized scores, medians.

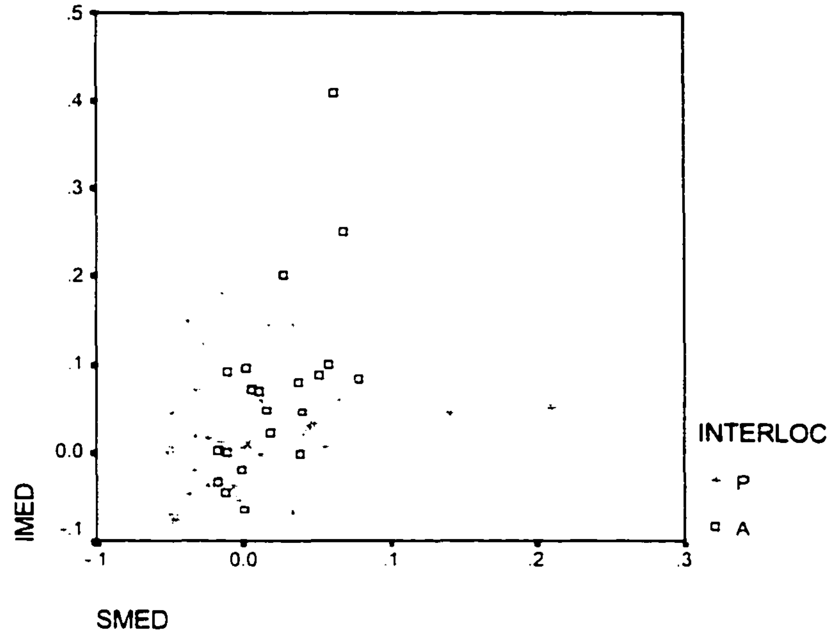


Figure 17. S9 with females, normalized scores, medians.

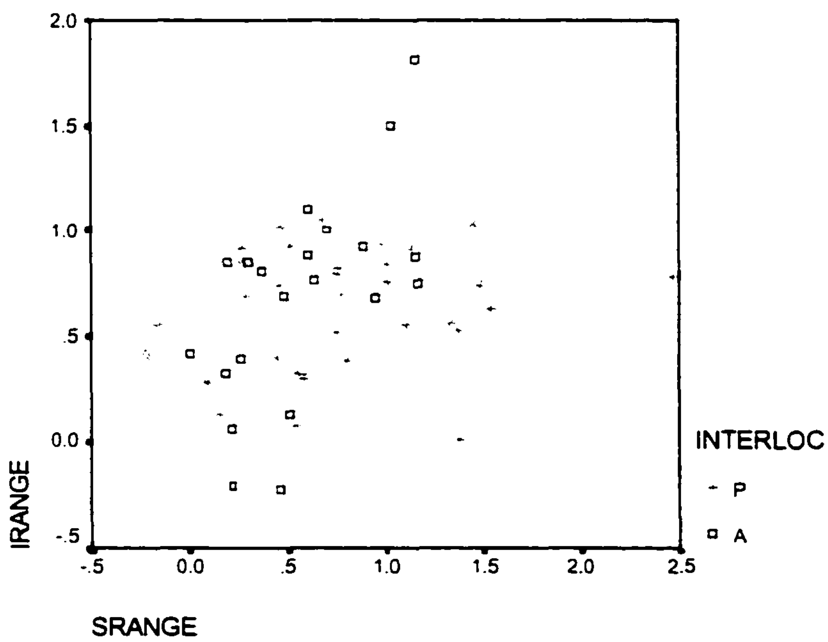


Figure 18. S9 with females, normalized scores, ranges.

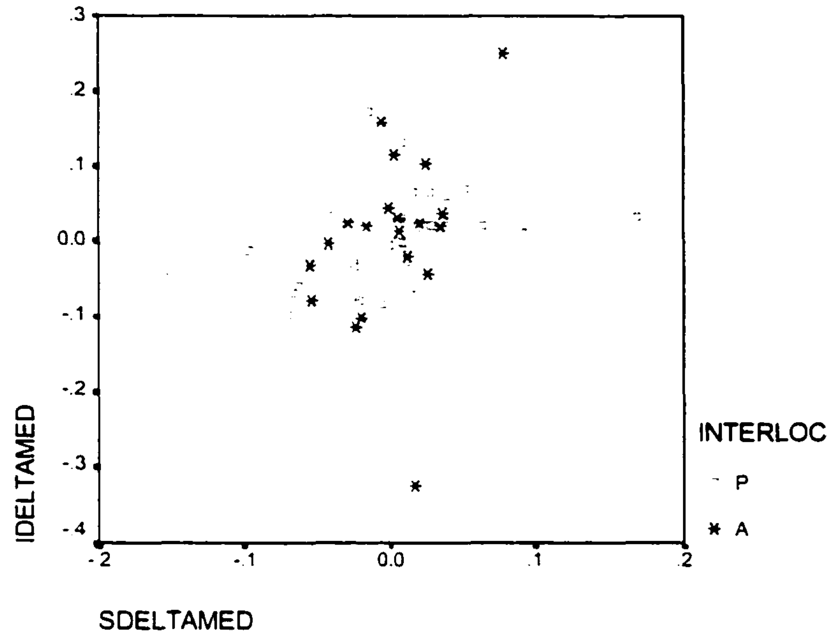


Figure 19. S9 with females, deltas, medians.

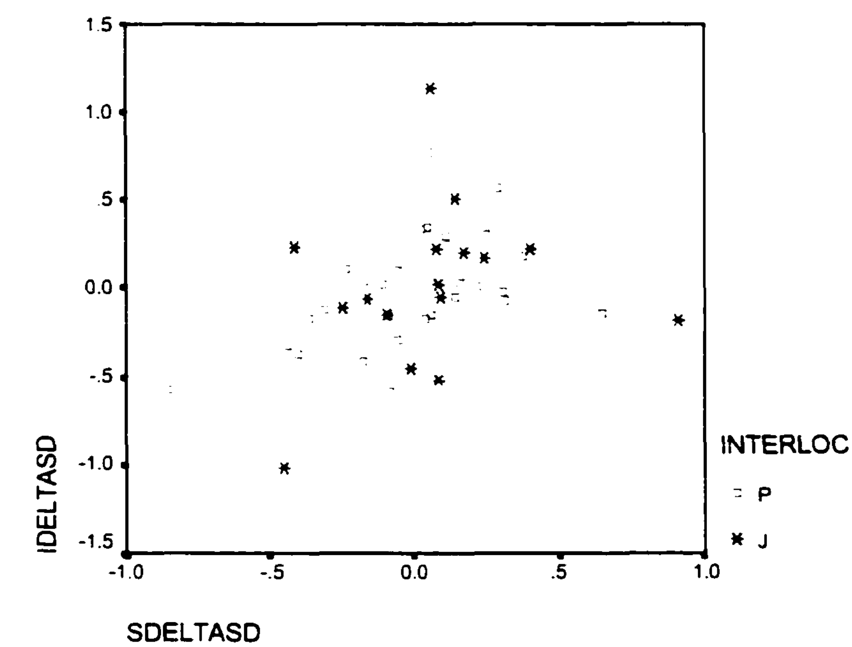


Figure 20. S10 with professors, deltas, SDs.

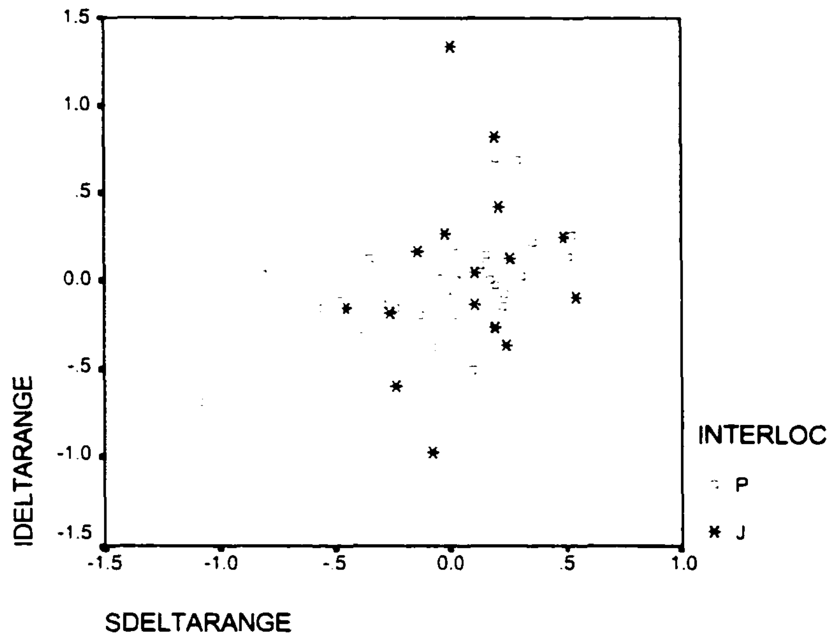


Figure 21. S10 with professors, deltas, ranges.

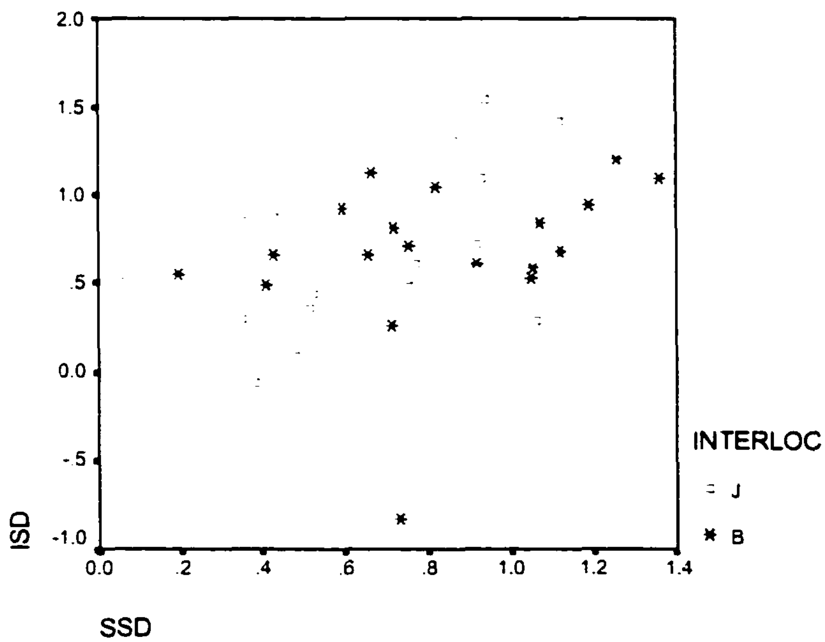


Figure 22. S10 with males, normalized scores, SDs.

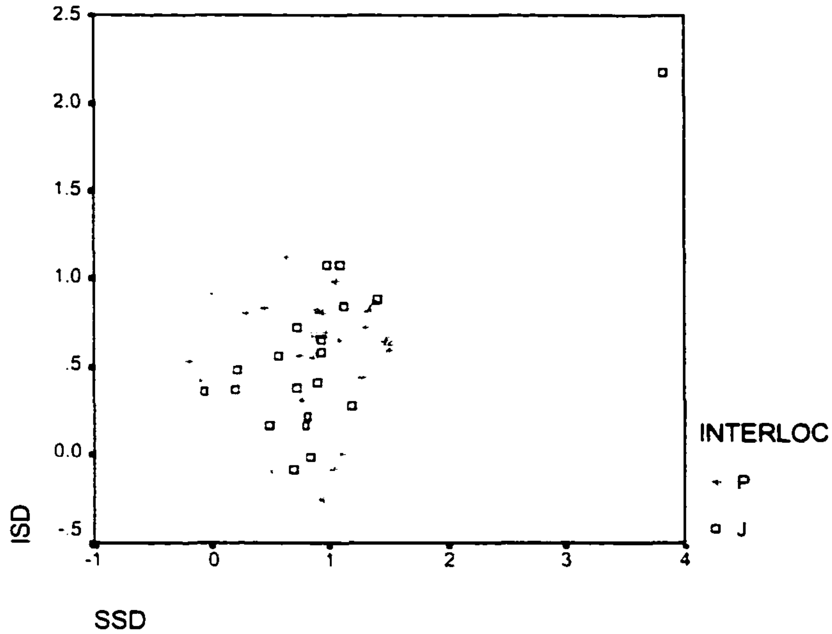


Figure 23. S11 with professors, normalized scores, SDs.

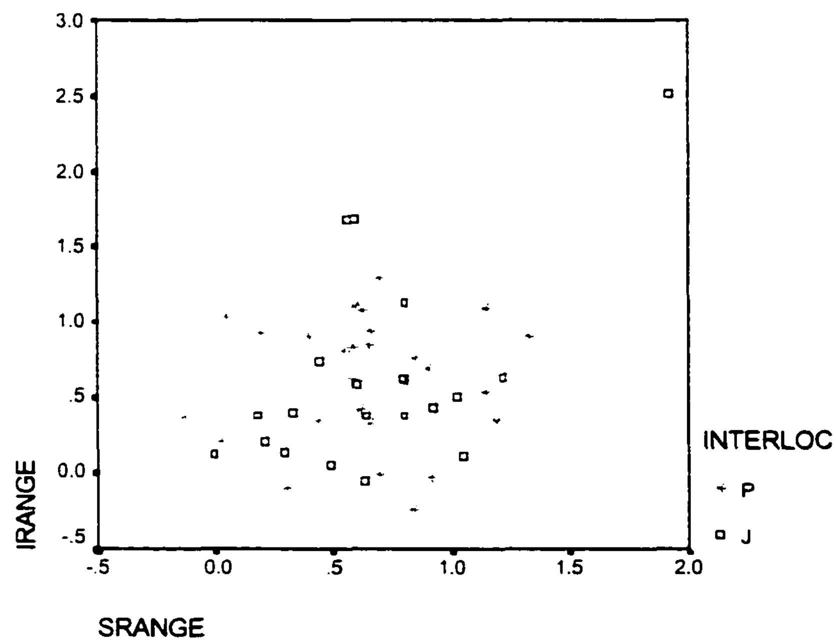


Figure 24. S11 with professors, normalized scores, ranges.

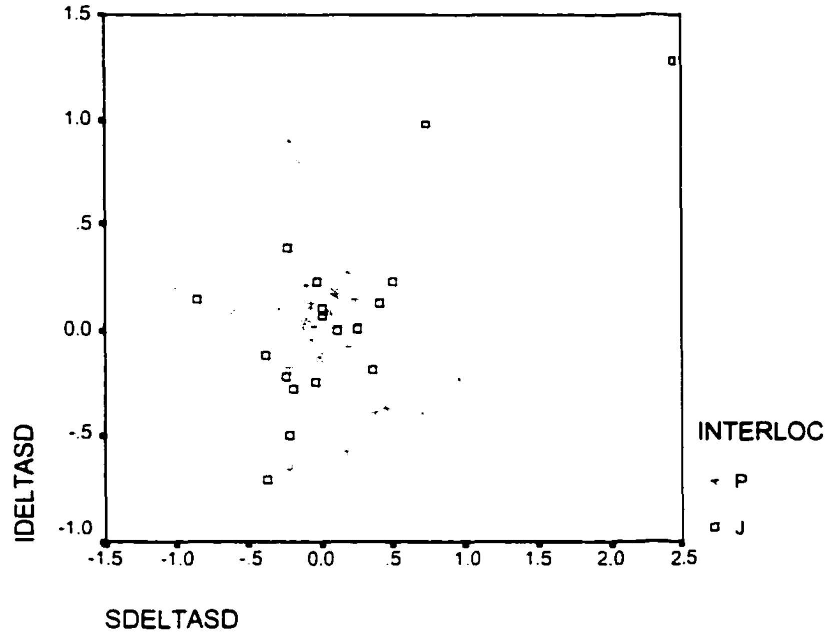


Figure 25. S11 with professors, deltas, SDs.

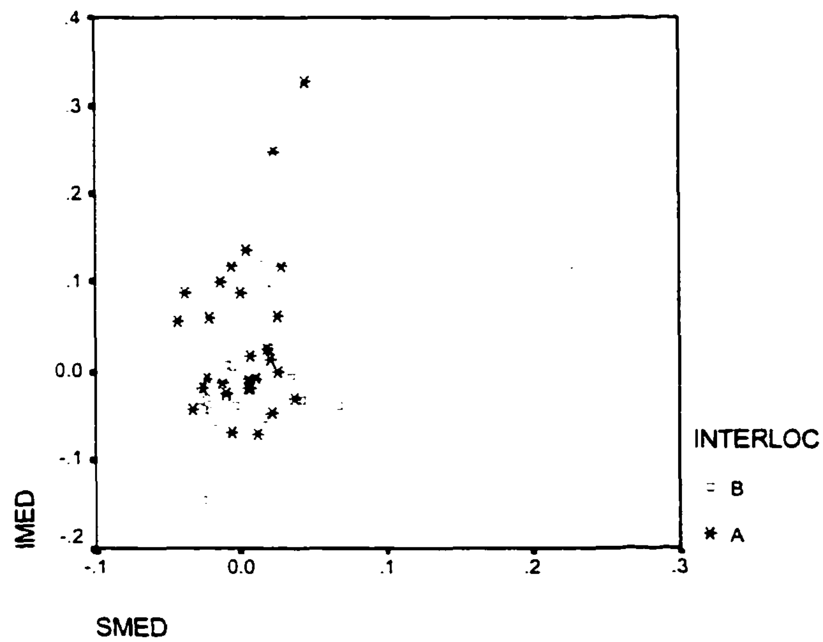


Figure 26. S11 with peers, normalized scores, medians.

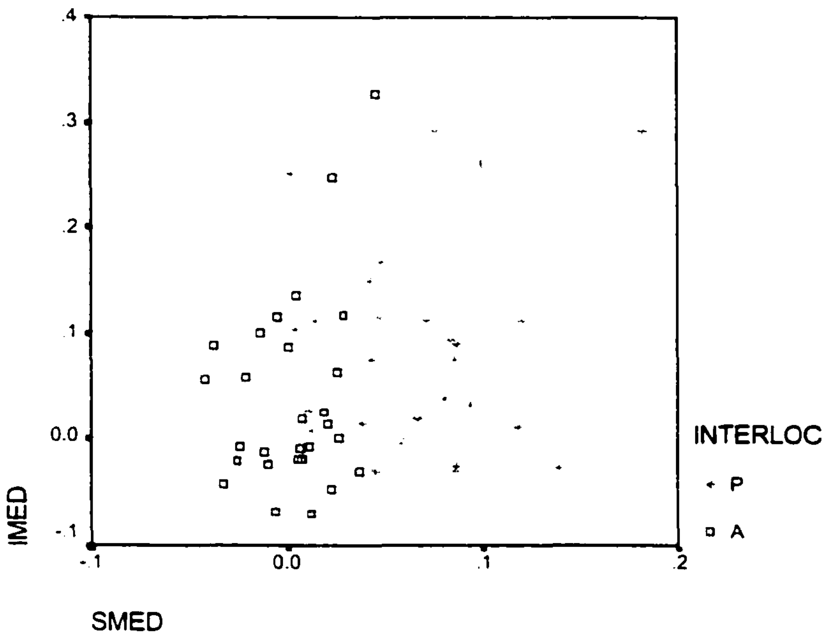


Figure 27. S11 with females, normalized scores, medians.

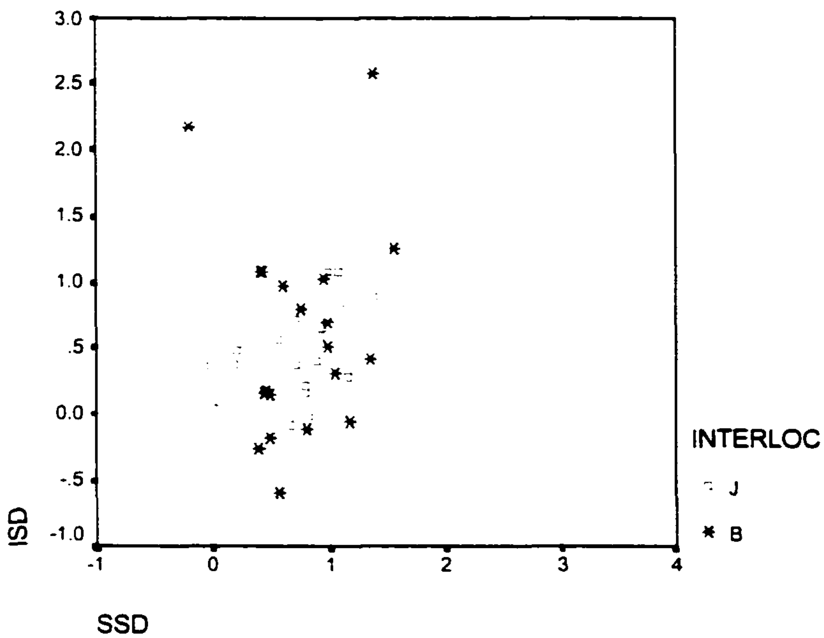


Figure 28. S11 with males, normalized scores, SDs.

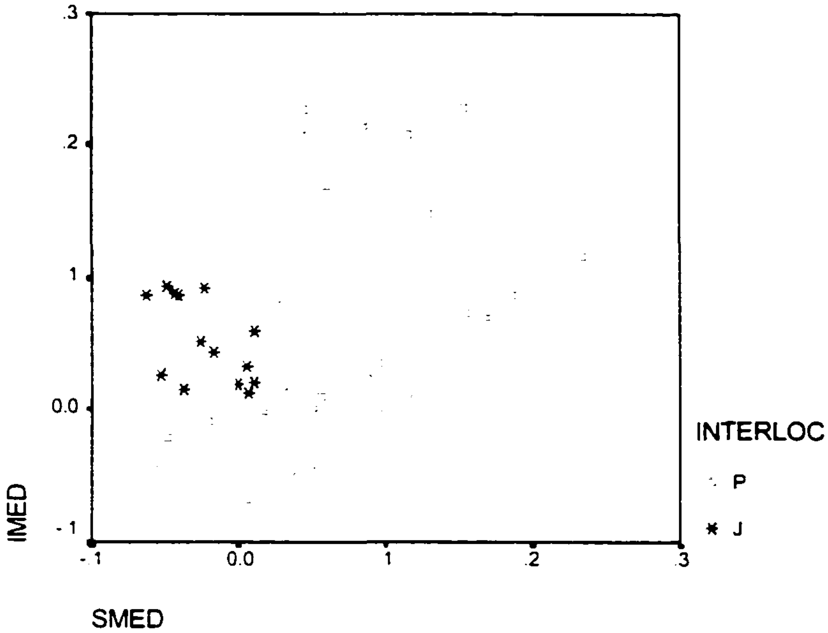


Figure 29. S12 with professors, normalized scores, medians.

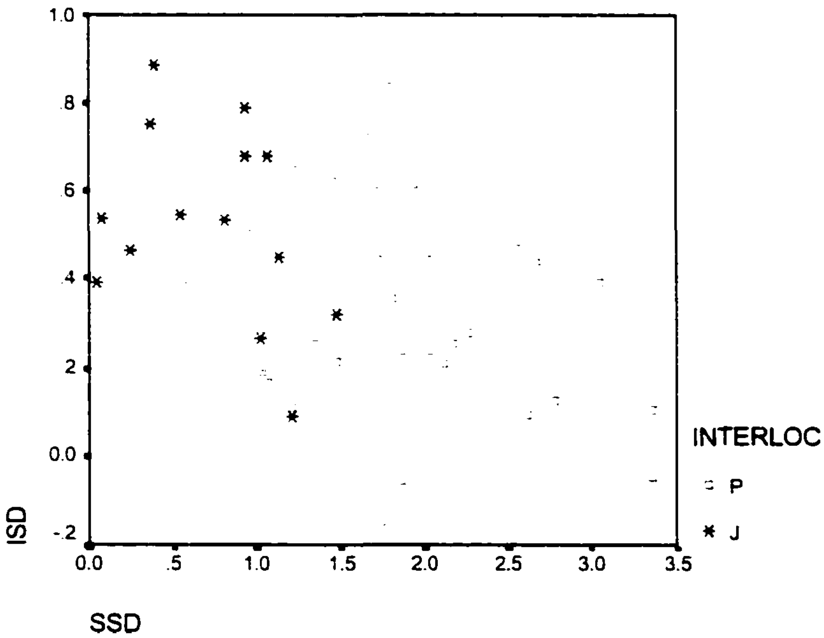


Figure 30. S12 with professors, normalized scores, SDs.

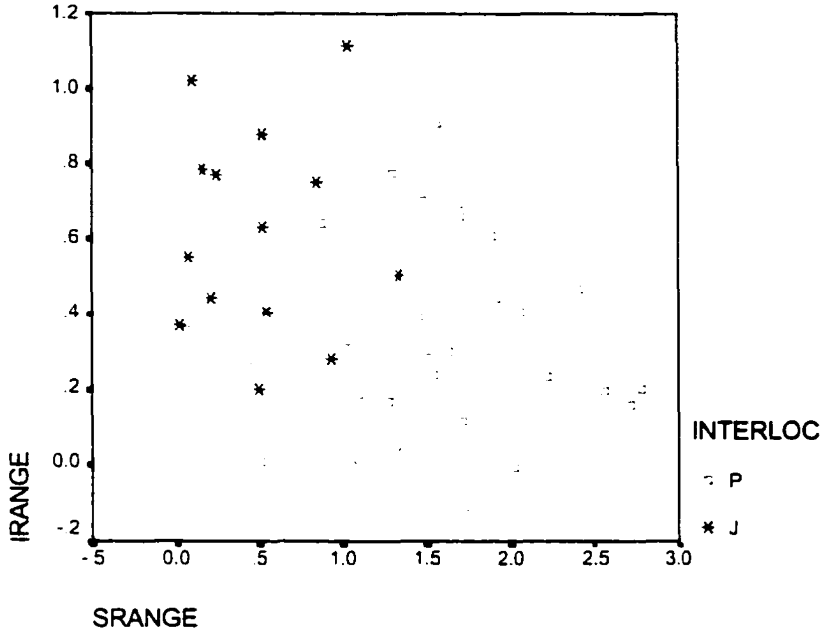


Figure 31. S12 with professors, normalized scores, ranges.

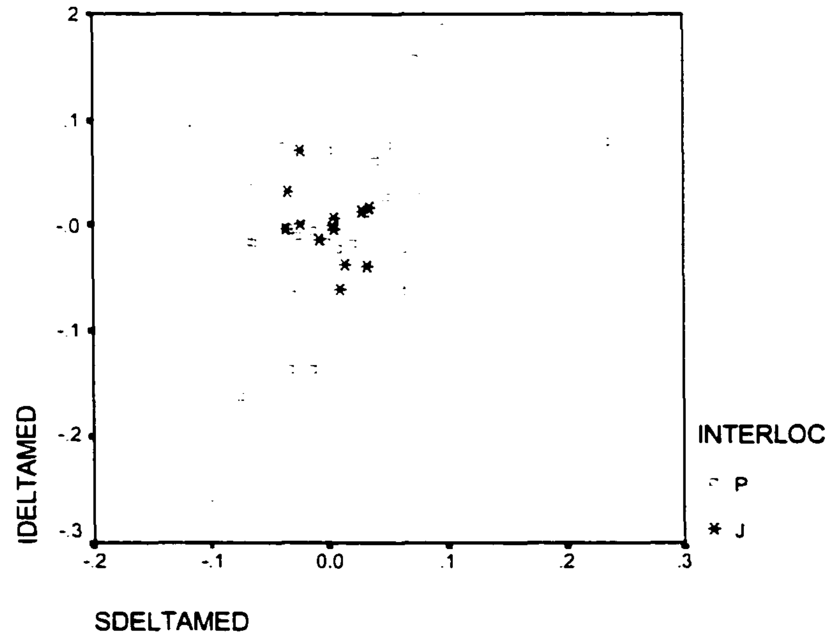
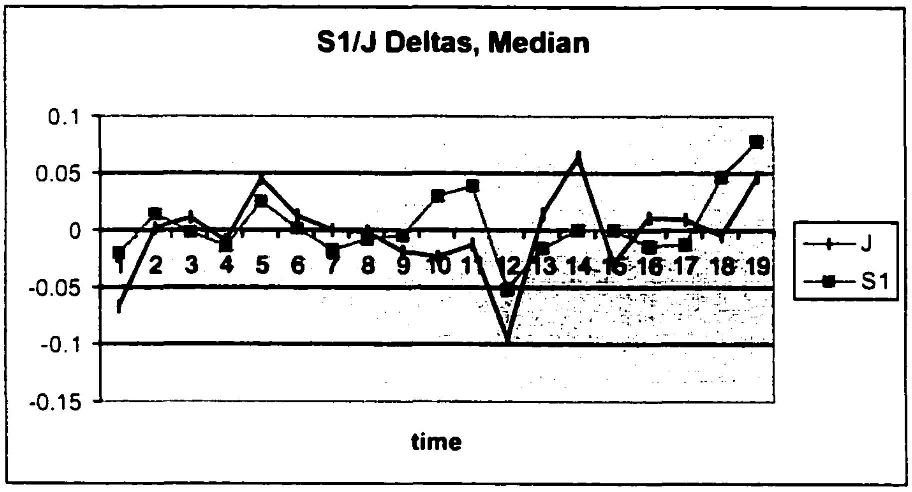


Figure 32. S12 with professors, deltas, medians.

Appendix 5: Significant Individual Interview Correlations, Complete Set of Graphs



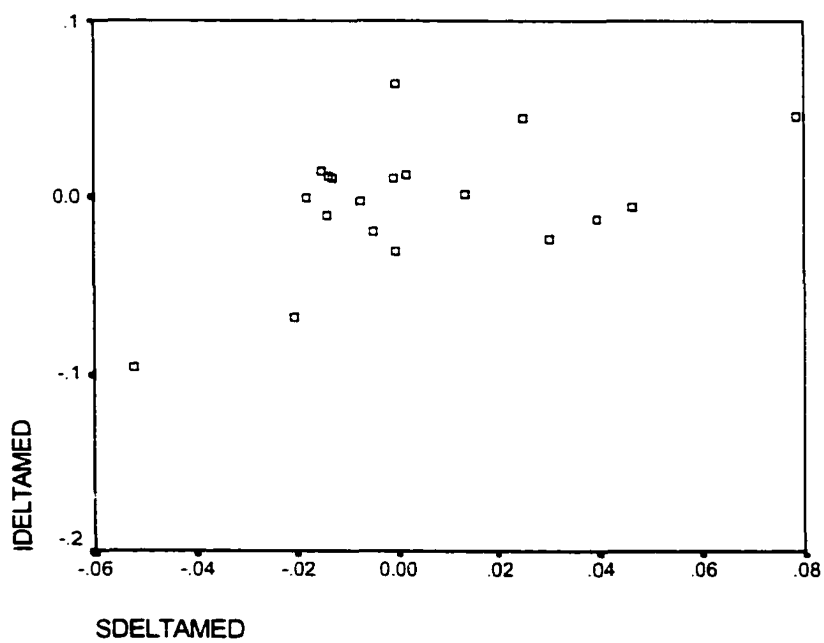
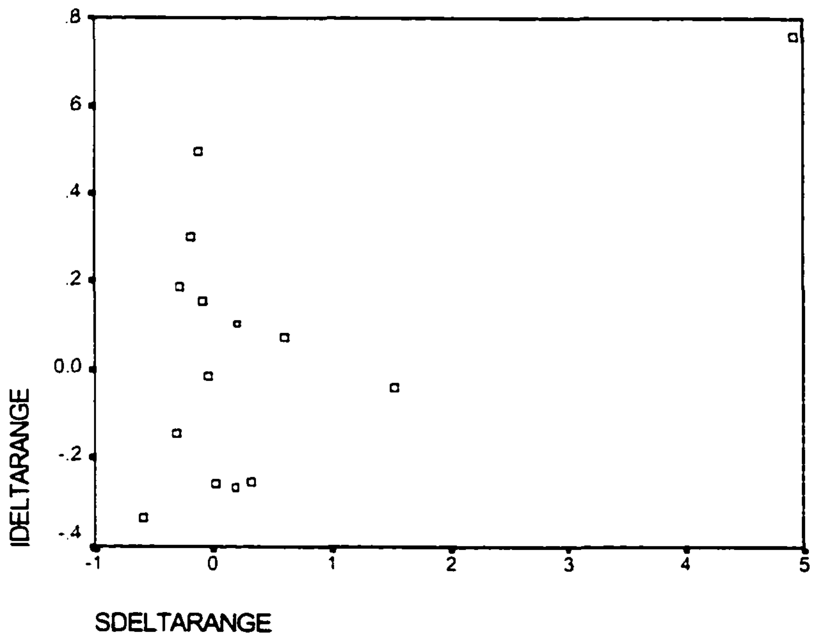


Figure 1. S1 with J, deltas, medians.



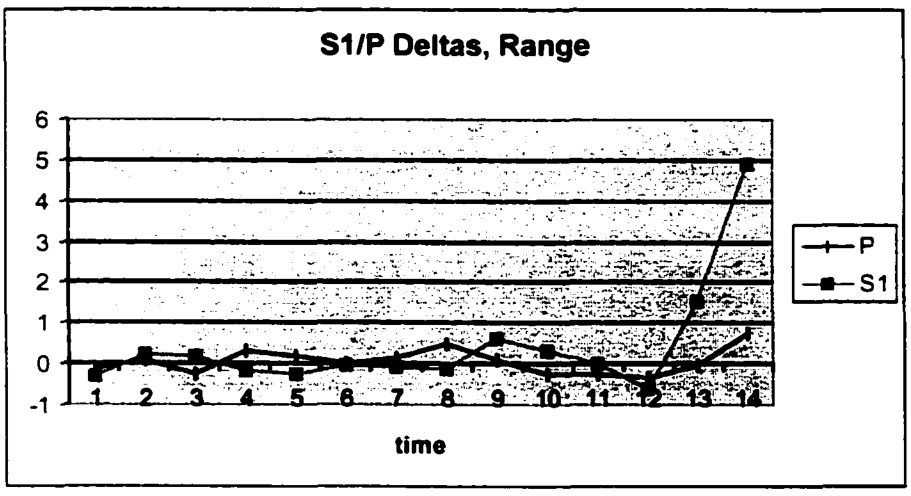
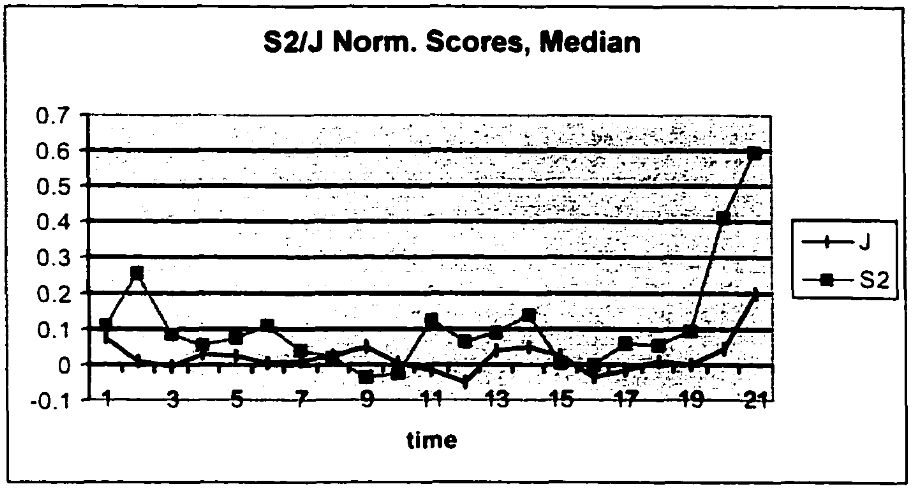


Figure 2. S1 with P, deltas, ranges.



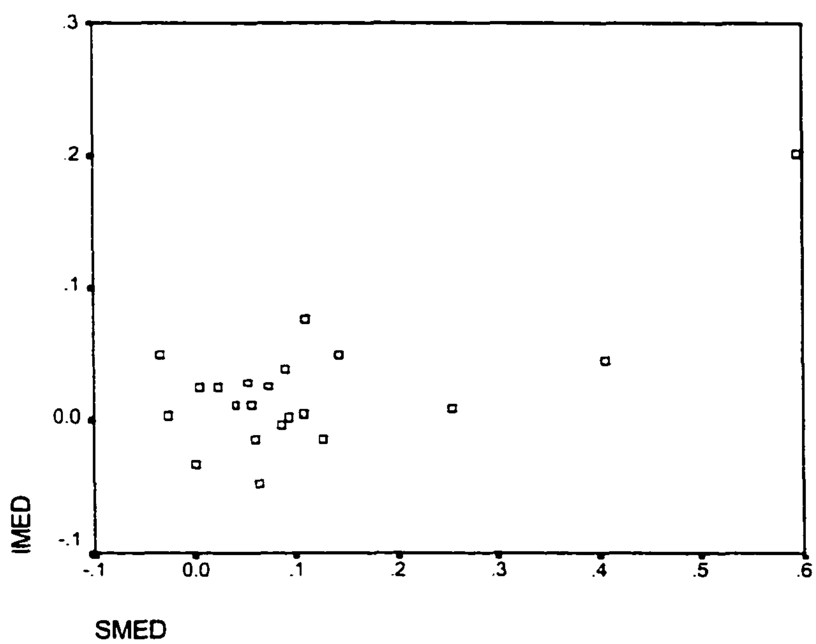
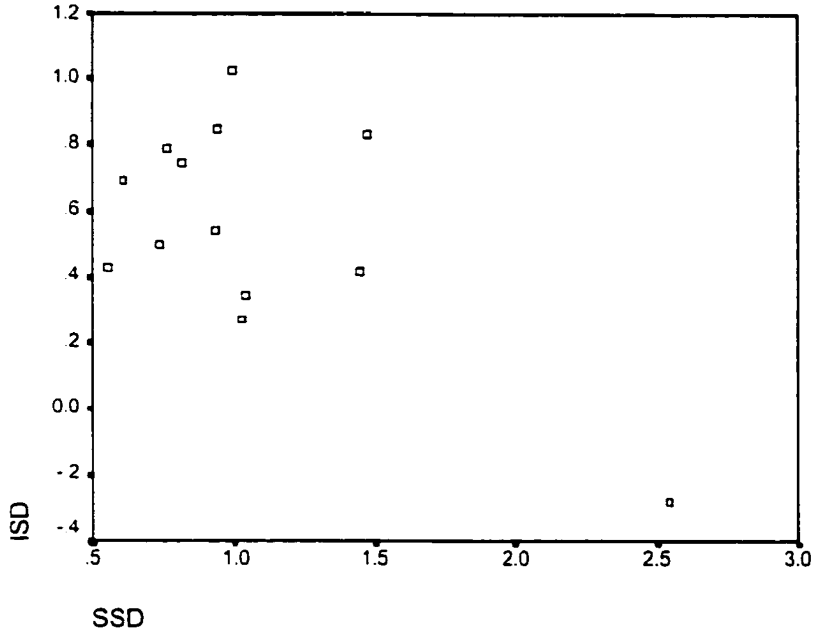


Figure 3. S2 with J, normalized scores, medians.



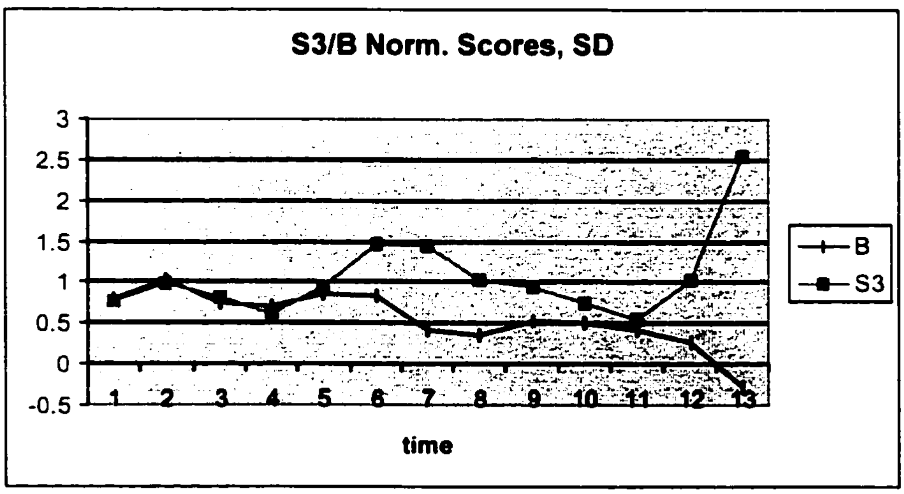
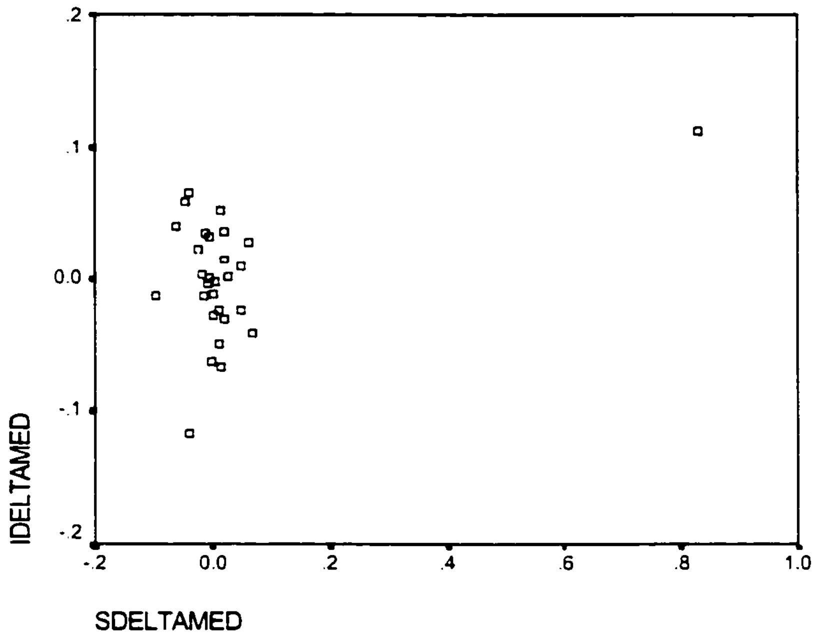


Figure 4. S3 with B, normalized scores, SDs.



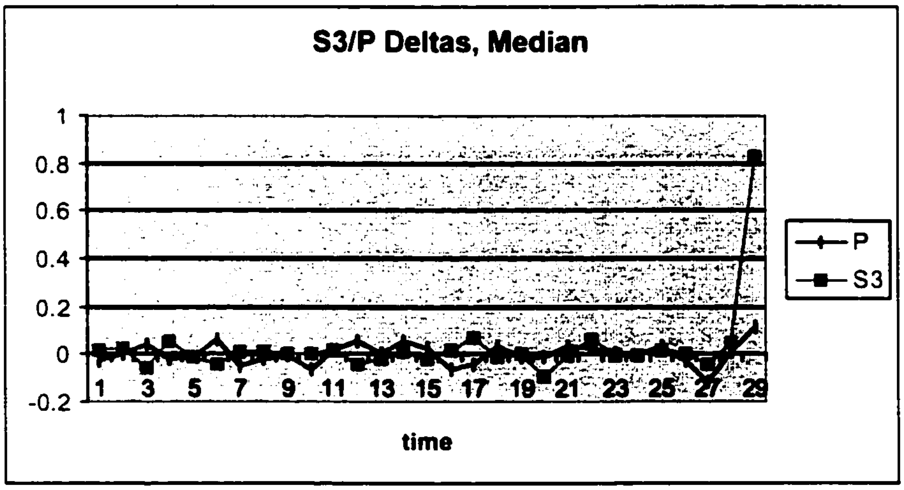


Figure 5. S3 with P, deltas, medians.

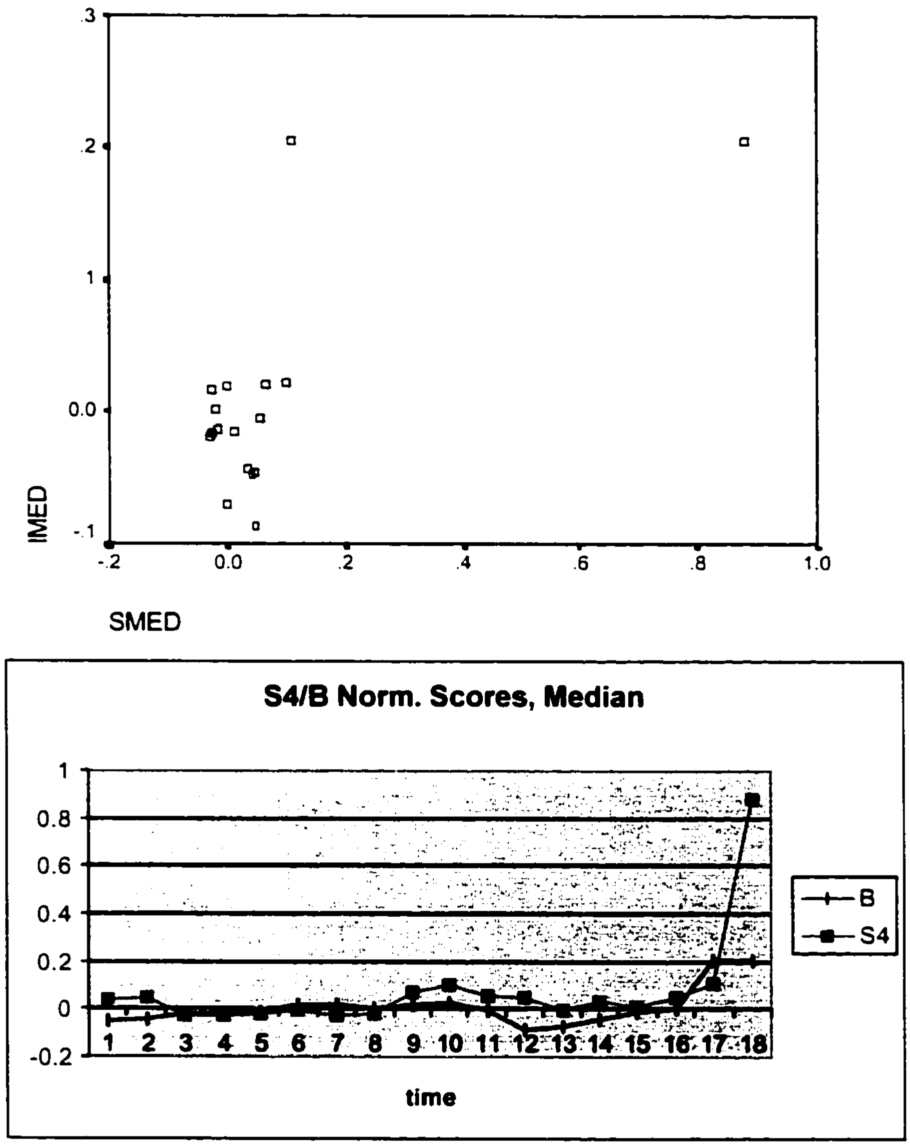


Figure 6. S4 with B, normalized scores, medians.

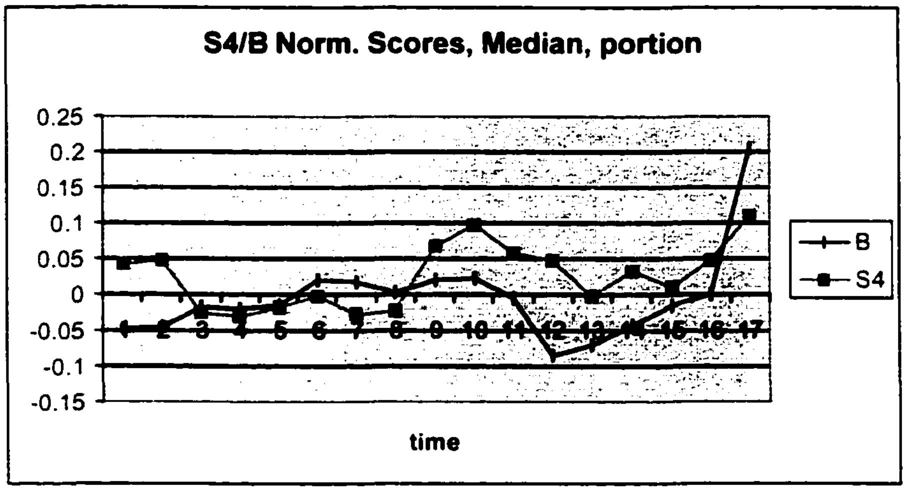
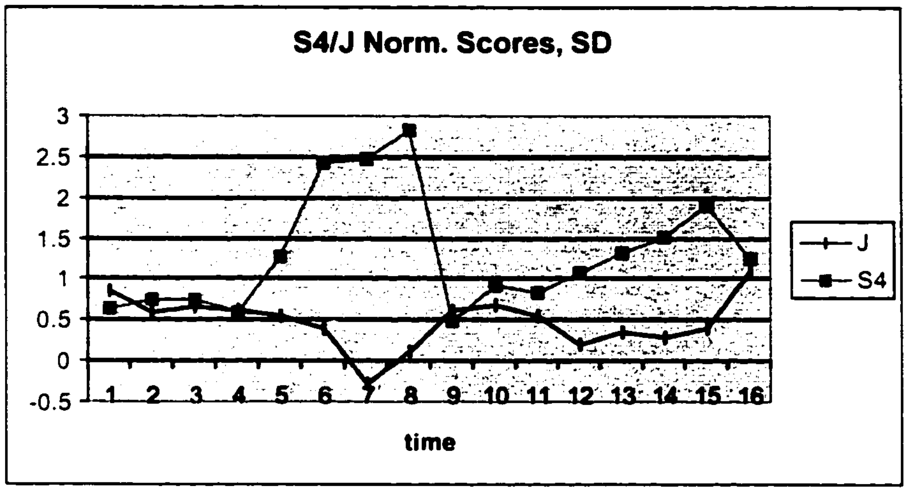


Figure 6b. S4 with B, normalized scores, medians, without possible outlier (data still correlates).



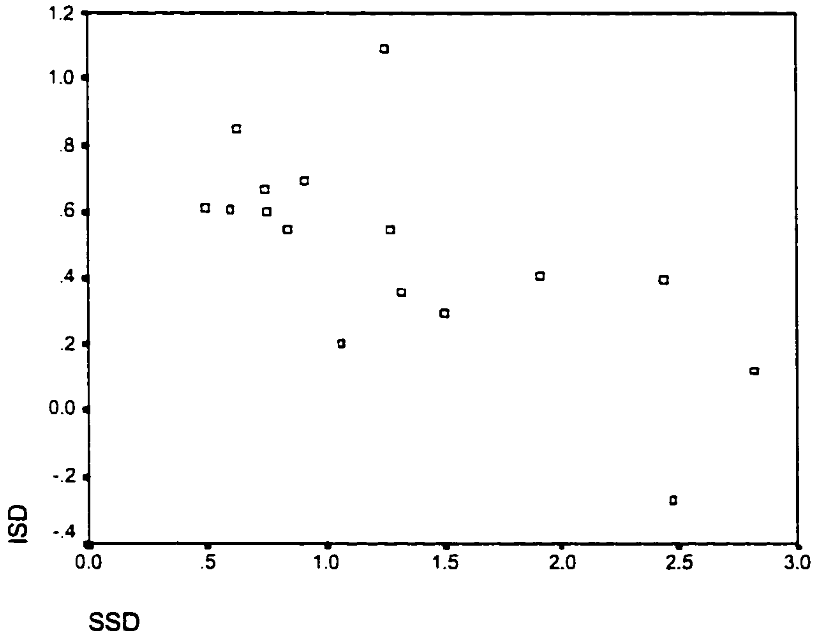
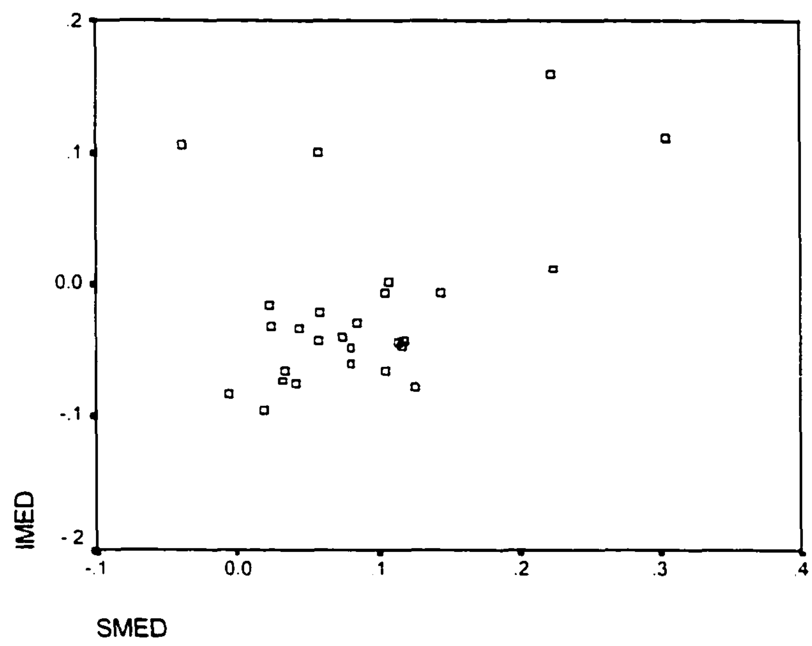


Figure 7. S4 with J, normalized scores, SDs.



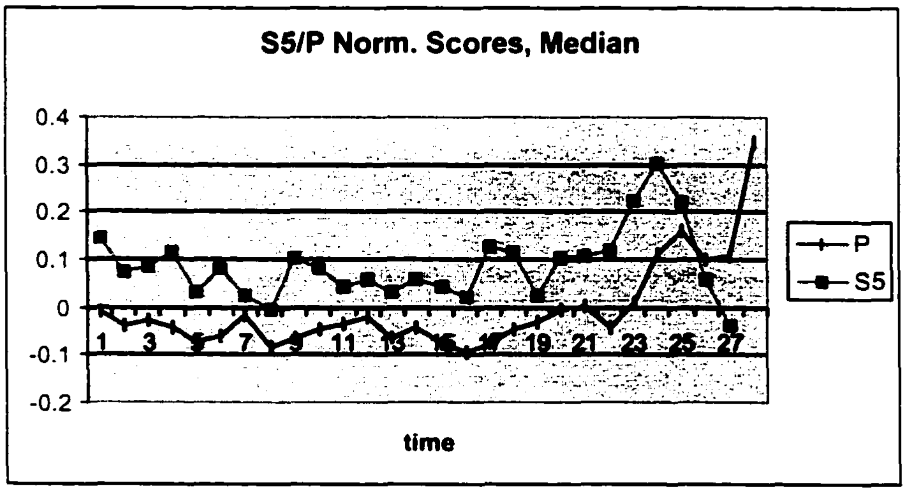
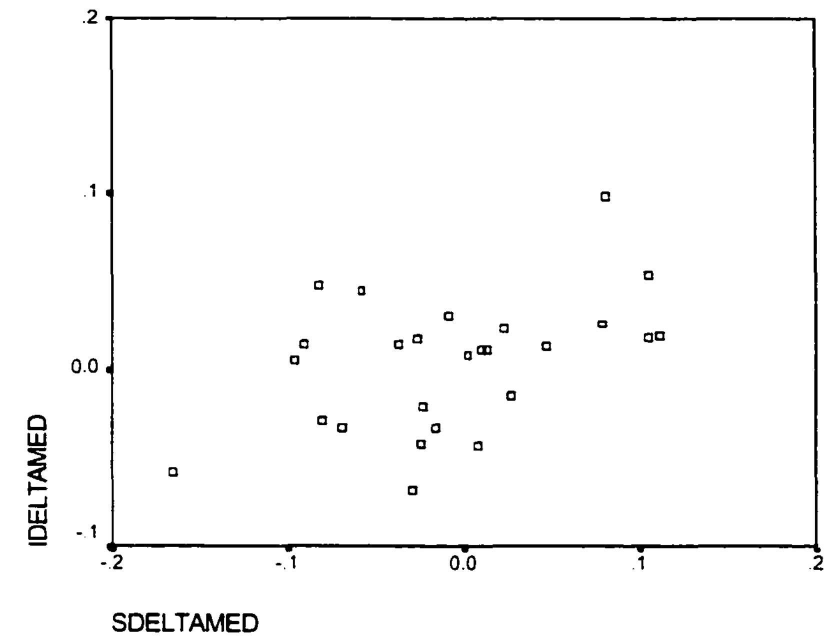


Figure 8. S5 with P, normalized scores, medians.



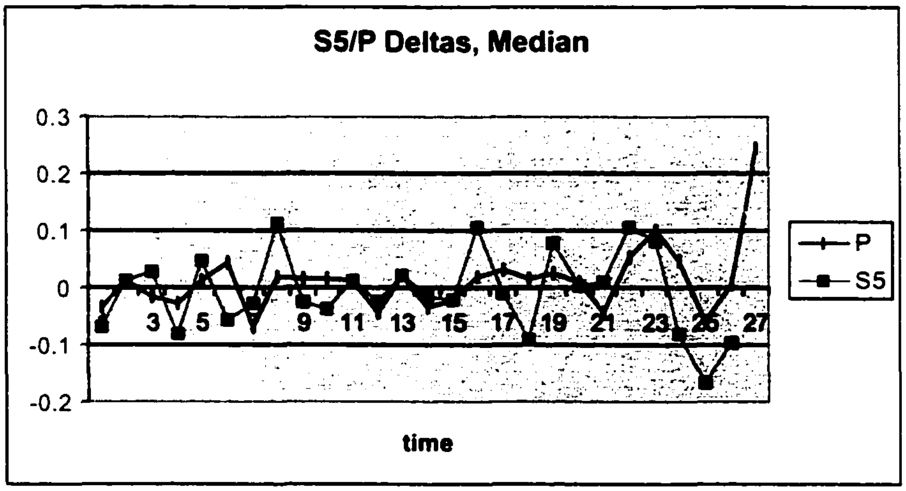
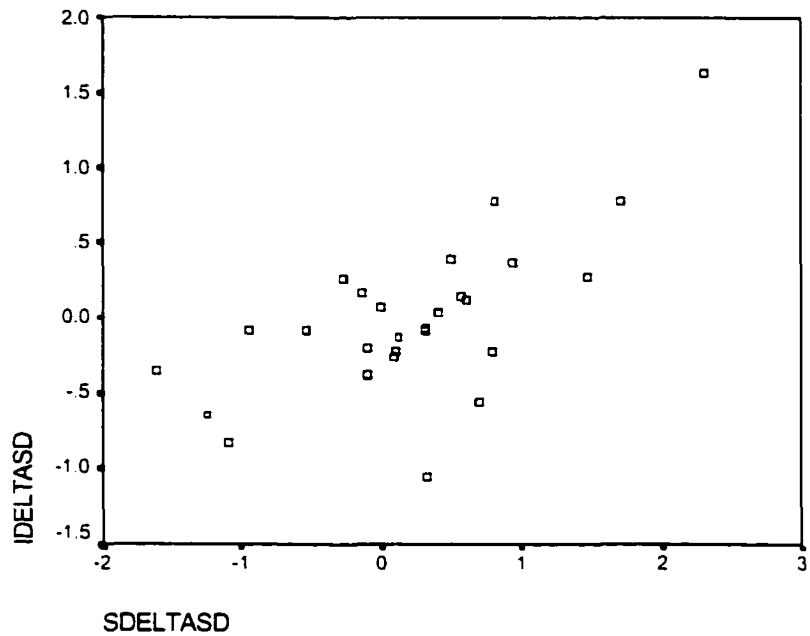


Figure 9. S5 with P, deltas, medians.



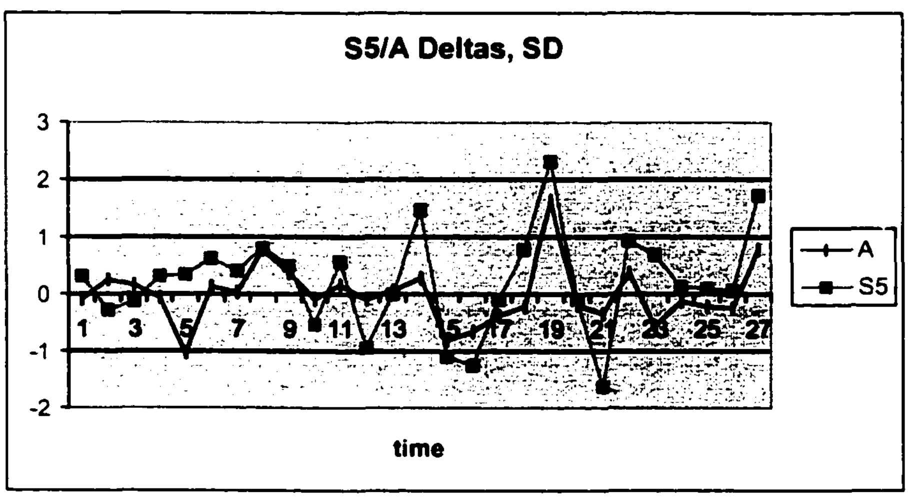
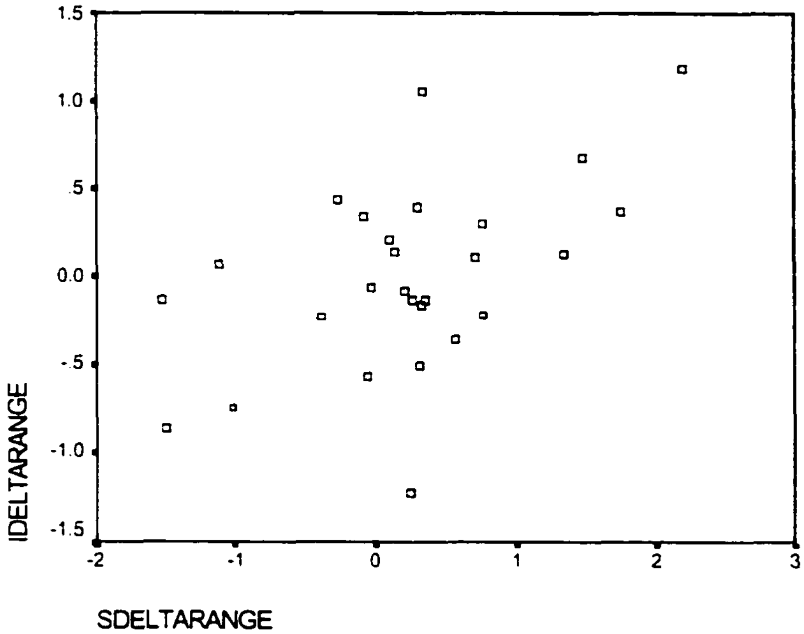


Figure 10. S5 with A, deltas, SDs.



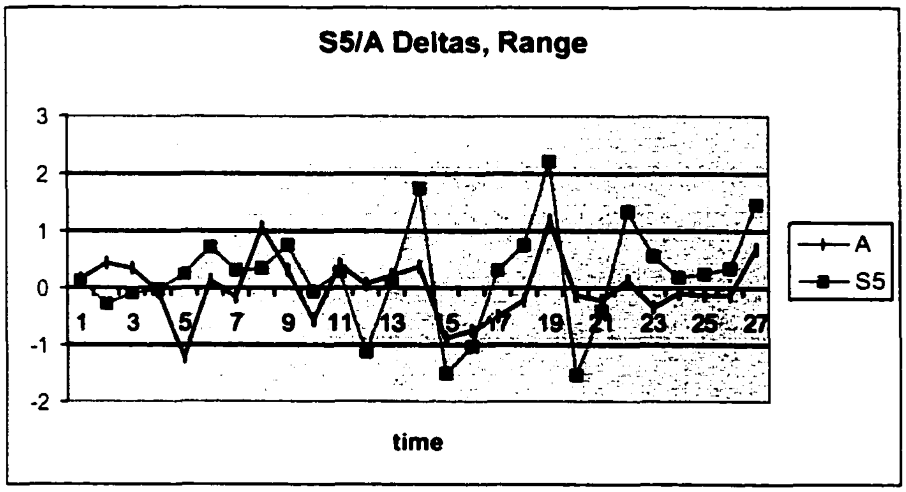
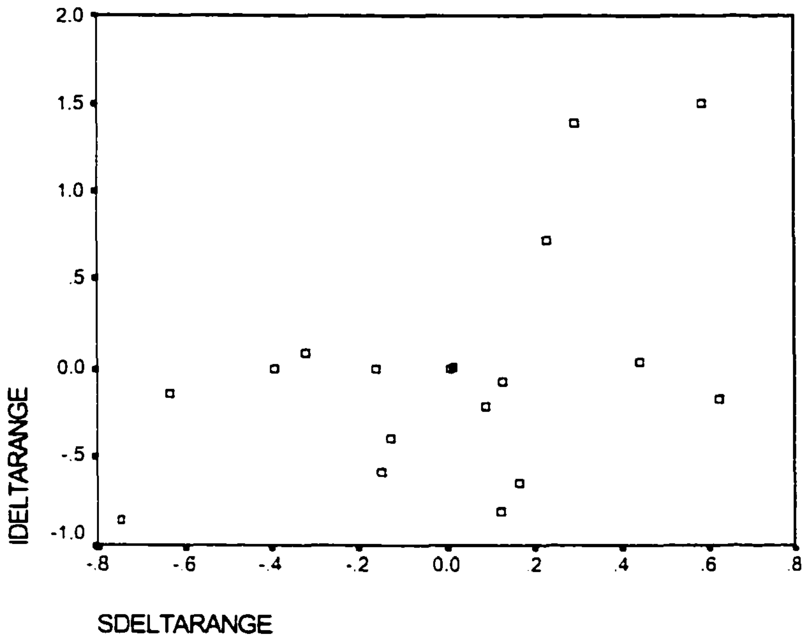


Figure 11. S5 with A, deltas, ranges.



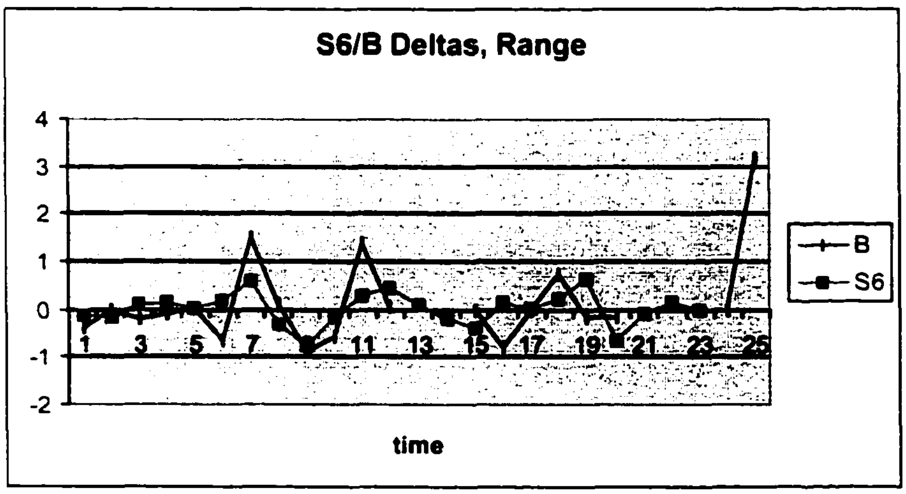
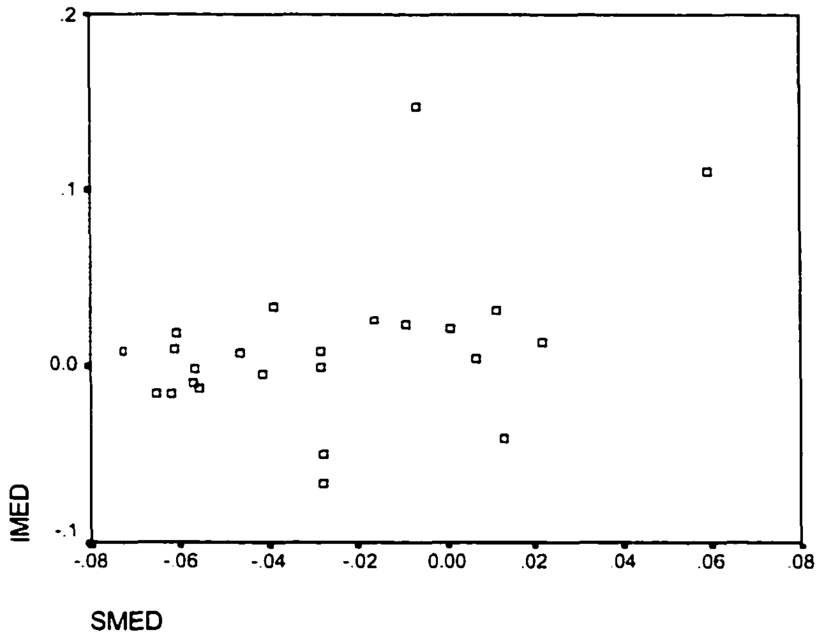


Figure 12. S6 with B, deltas, ranges.



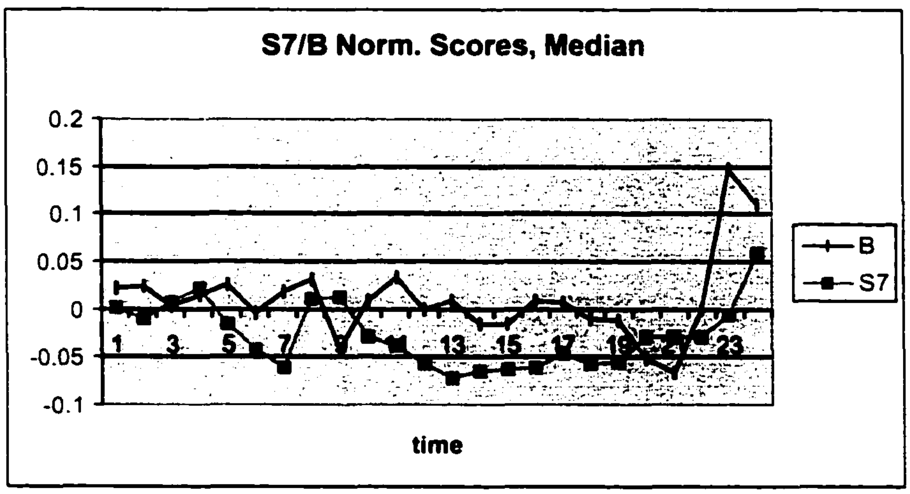
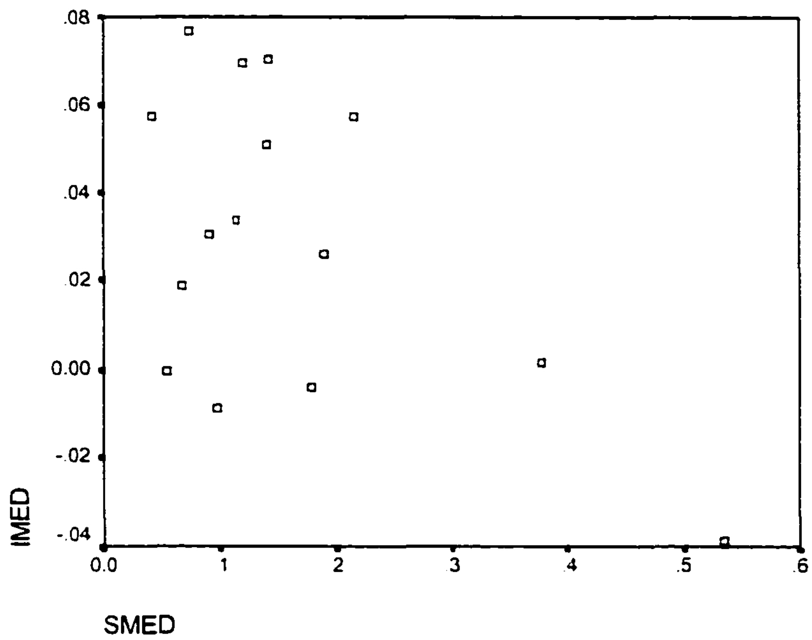


Figure 13. S7 with B, normalized scores, medians.



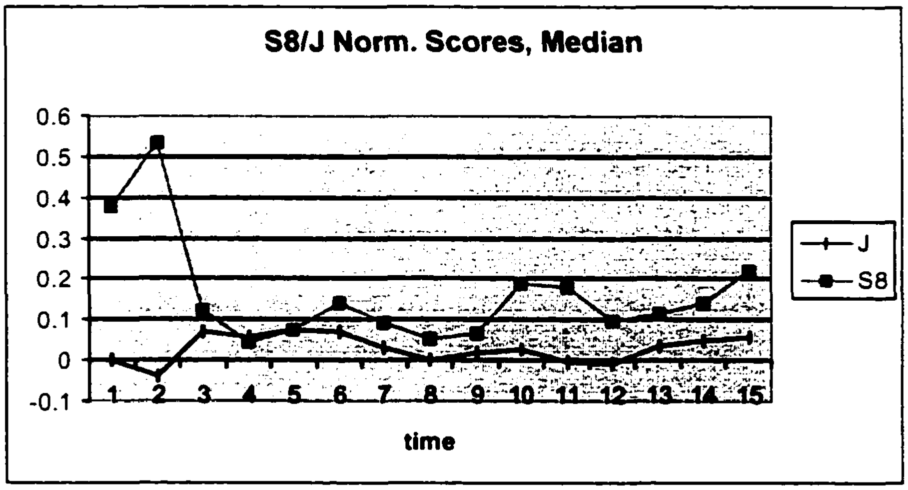
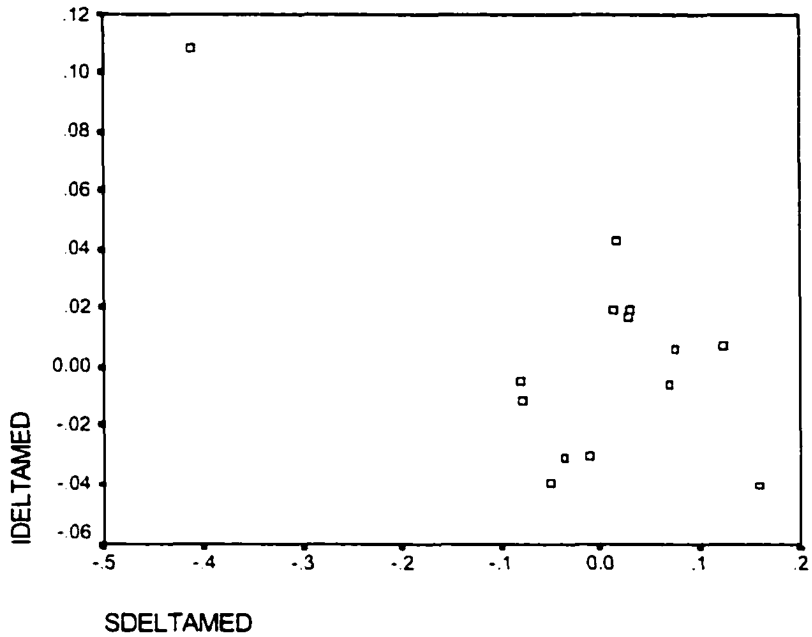


Figure 14. S8 with J, normalized scores, medians.



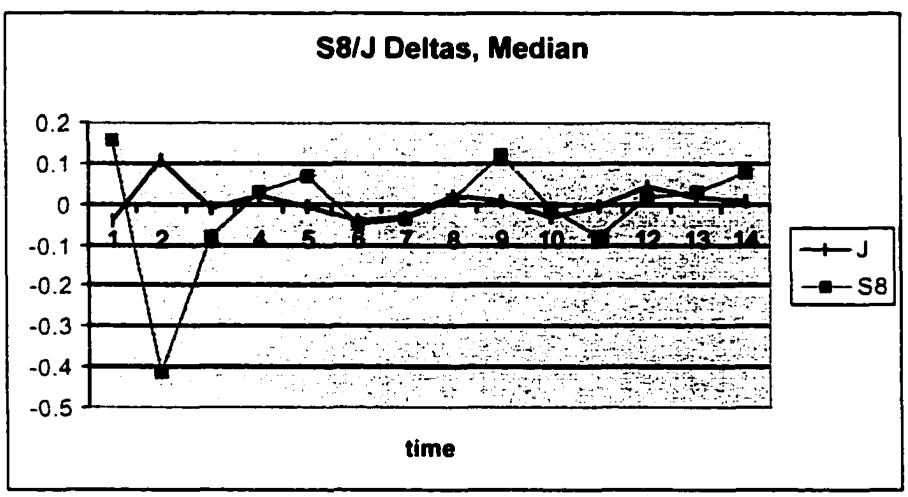
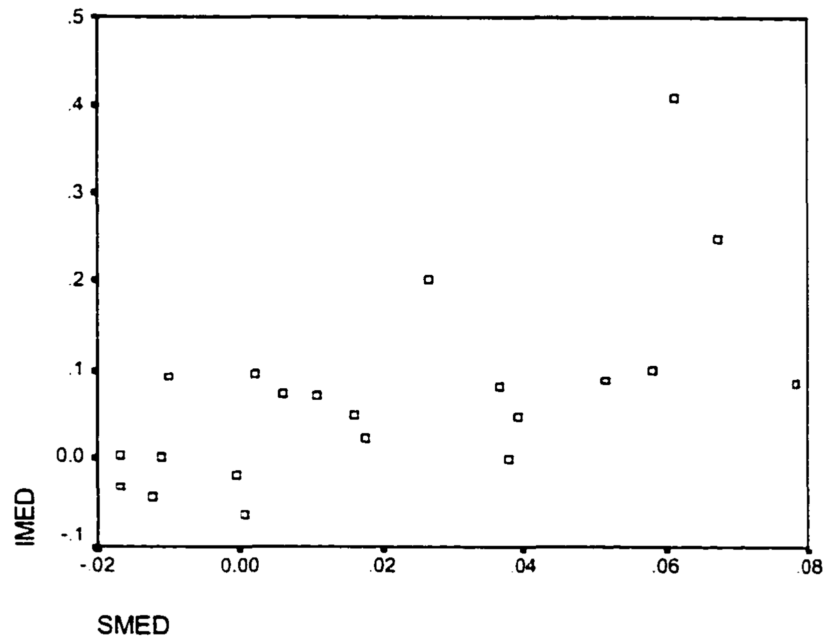


Figure 15. S8 with J, deltas, medians.



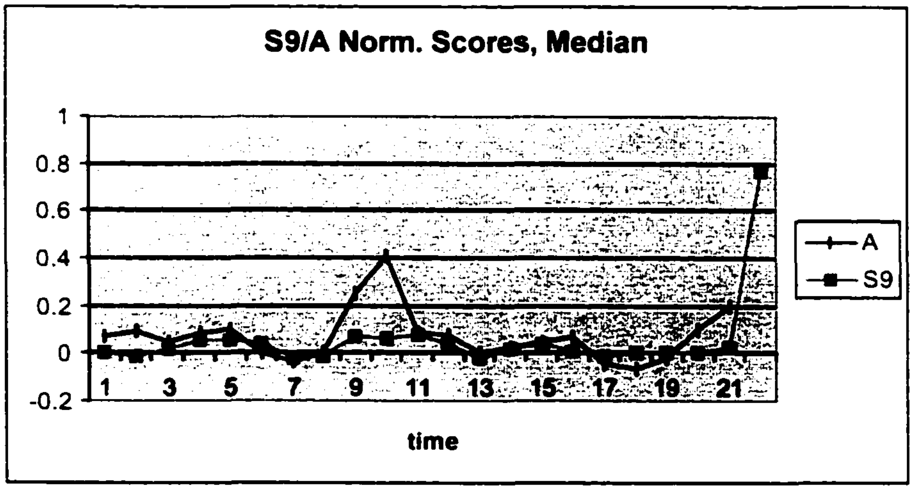
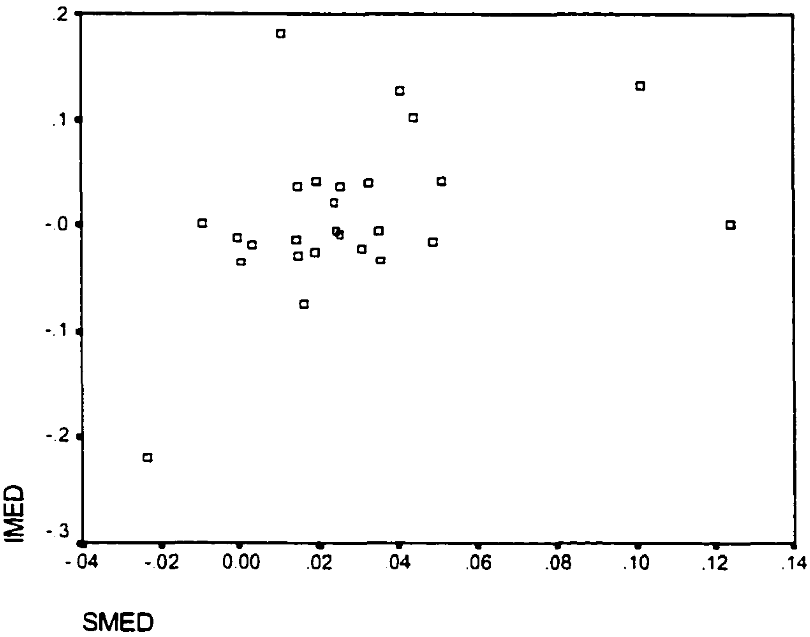


Figure 16. S9 with A, normalized scores, medians.



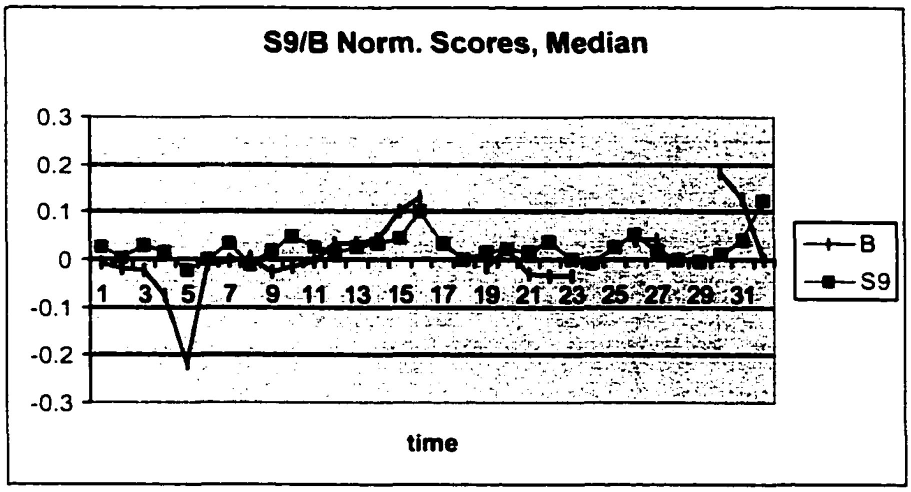
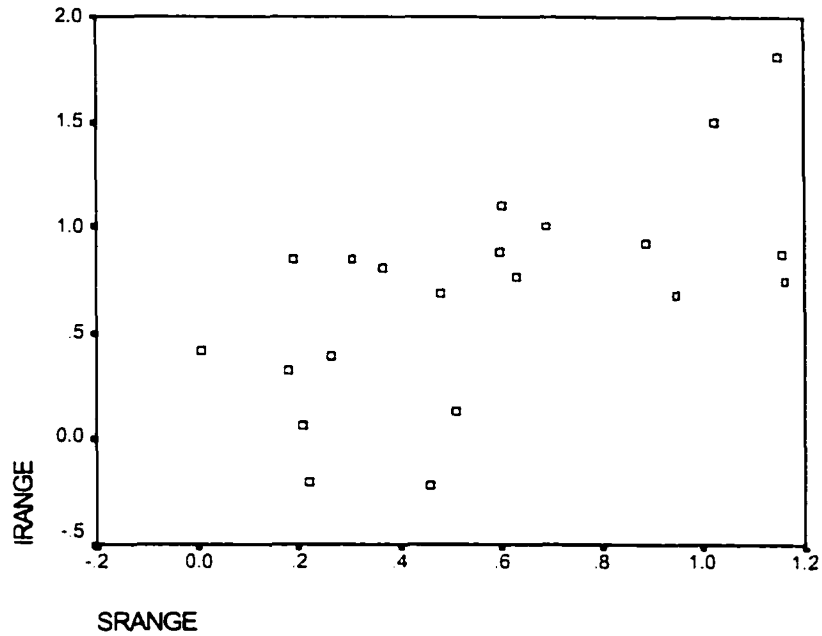


Figure 17. S9 with B, normalized scores, medians.



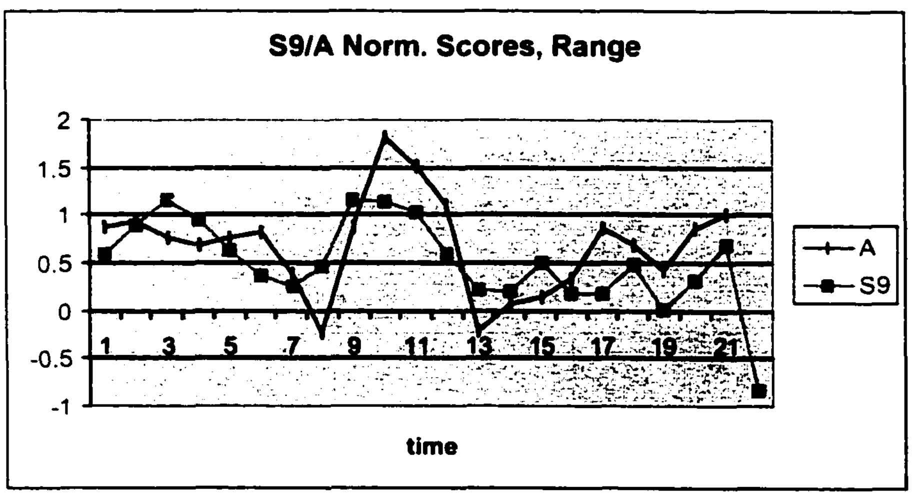
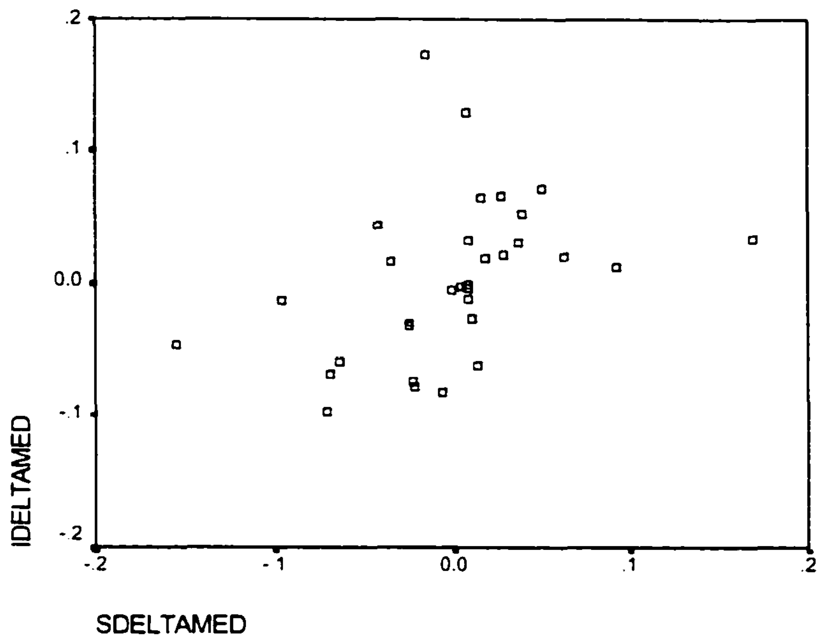


Figure 18. S9 with A, normalized scores, ranges.



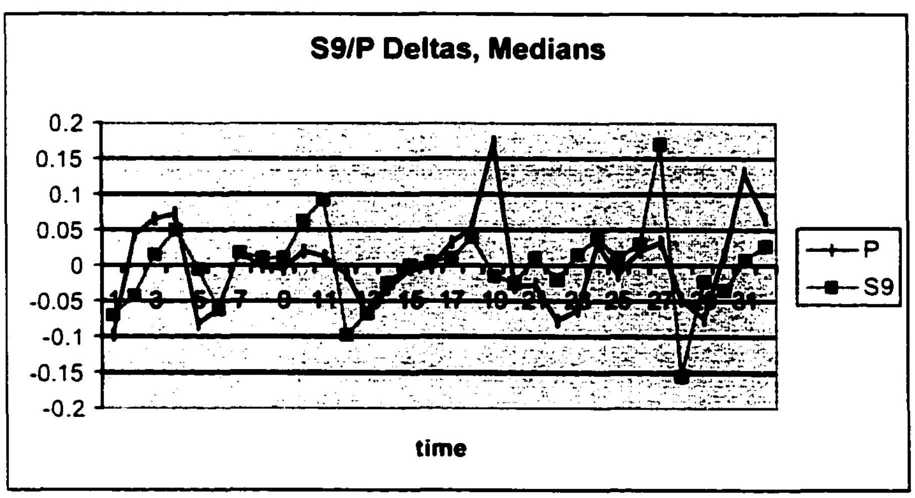
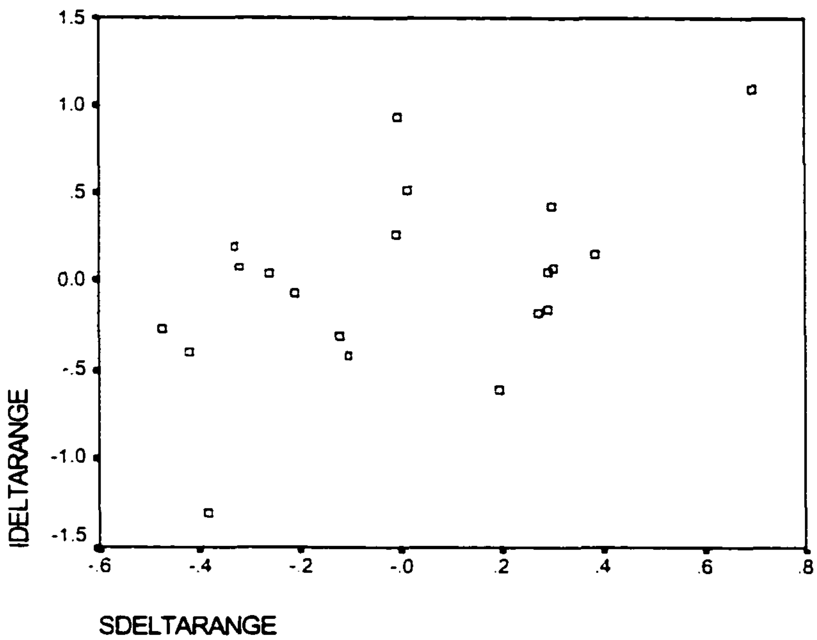


Figure 19. S9 with P, deltas, medians.



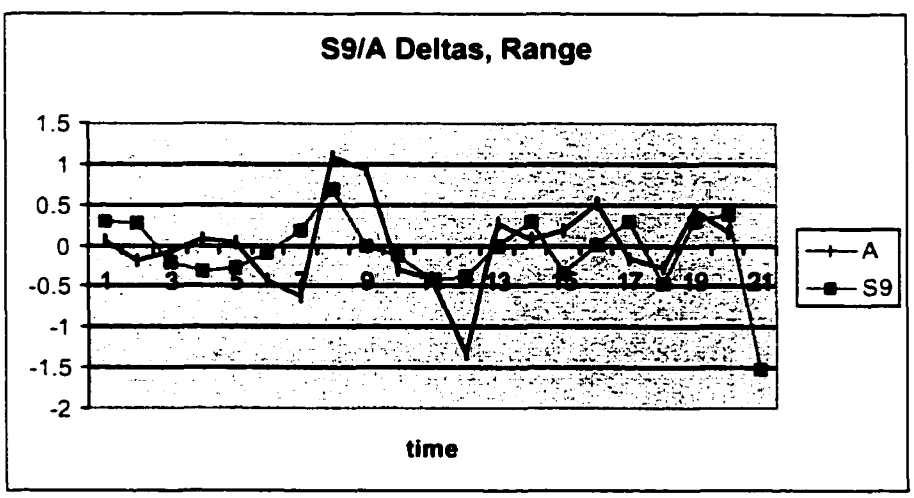
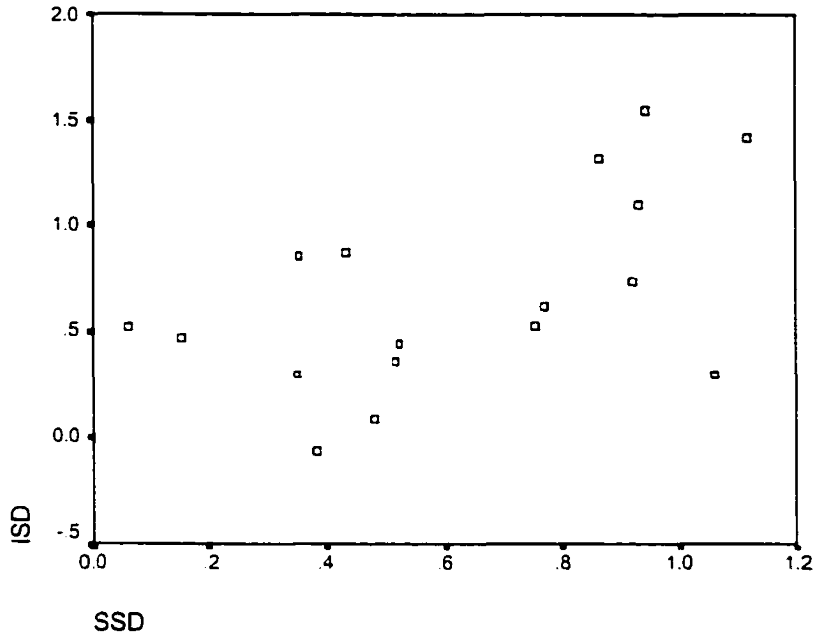


Figure 20. S9 with A, deltas, ranges.



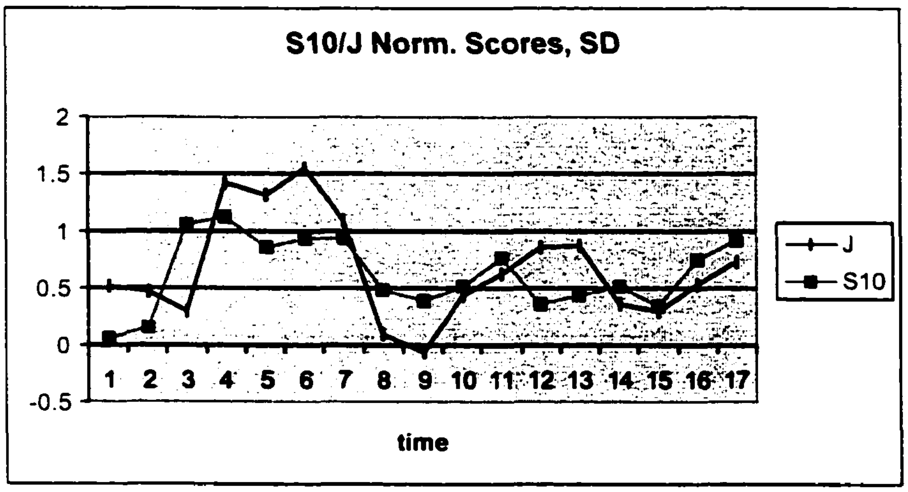
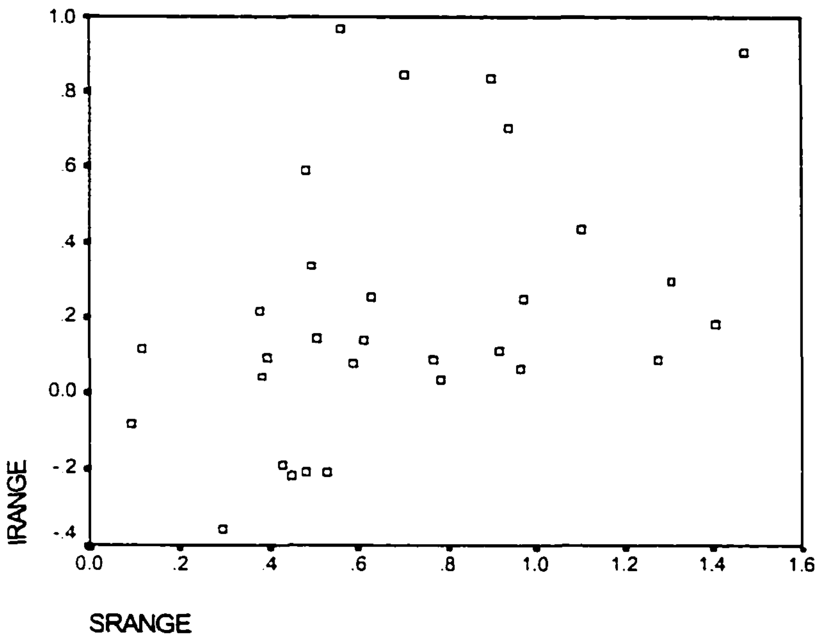


Figure 21. S10 with J, normalized scores, SDs.



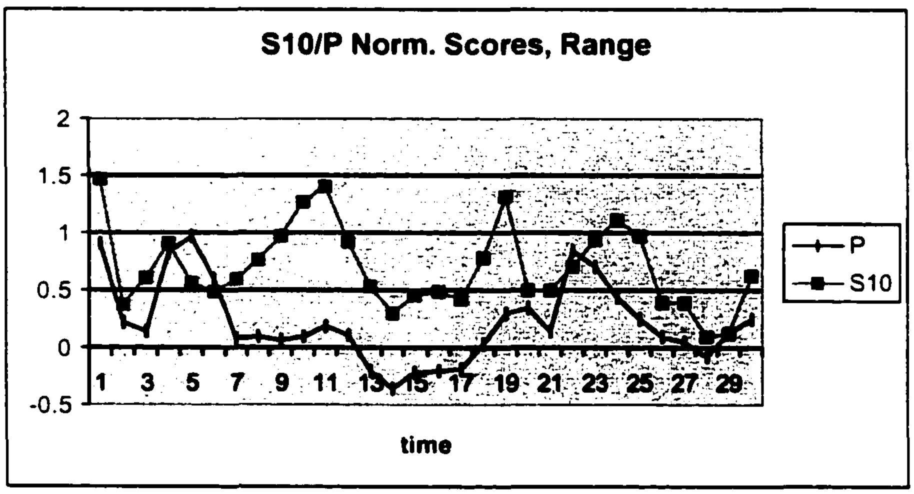
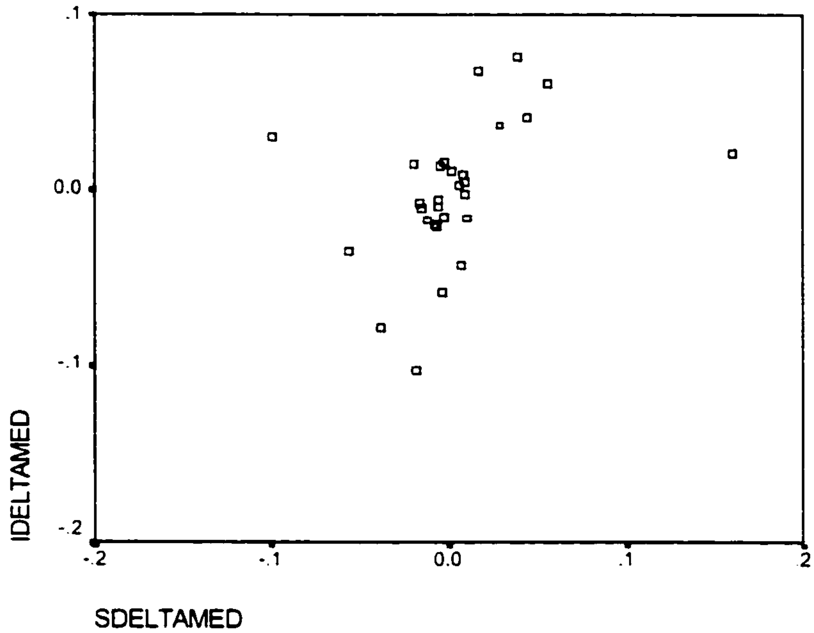


Figure 22. S10 with P, normalized scores, ranges.



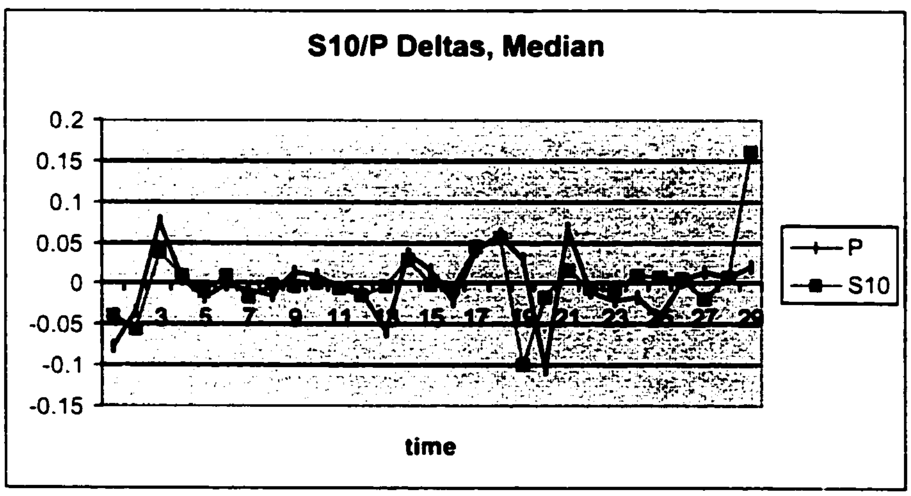
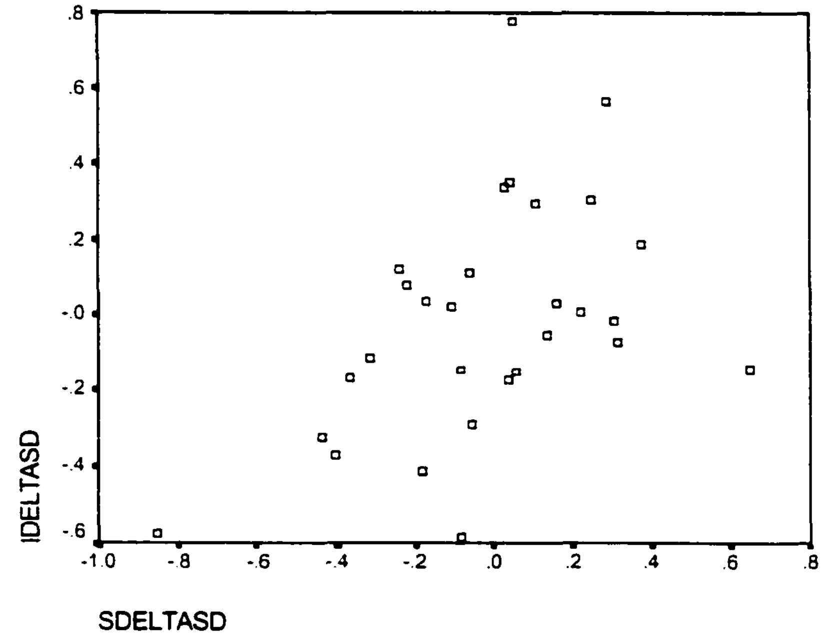


Figure 23. S10 with P, deltas, medians.



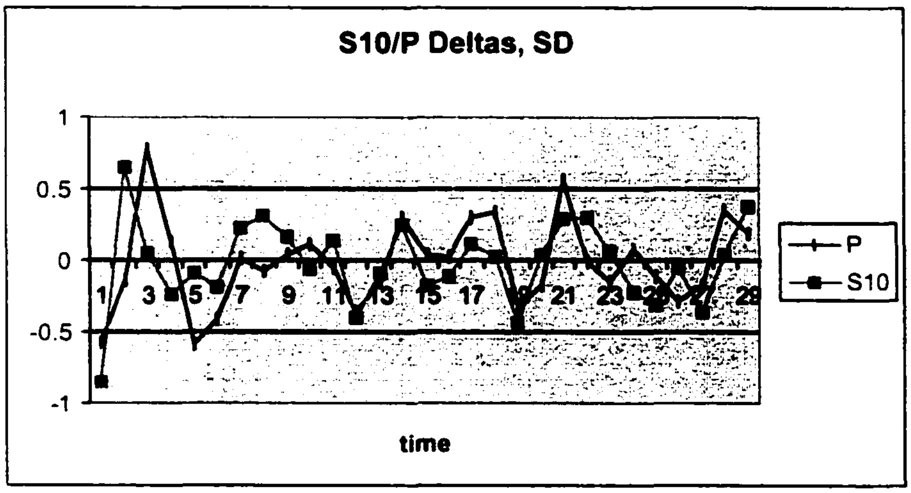
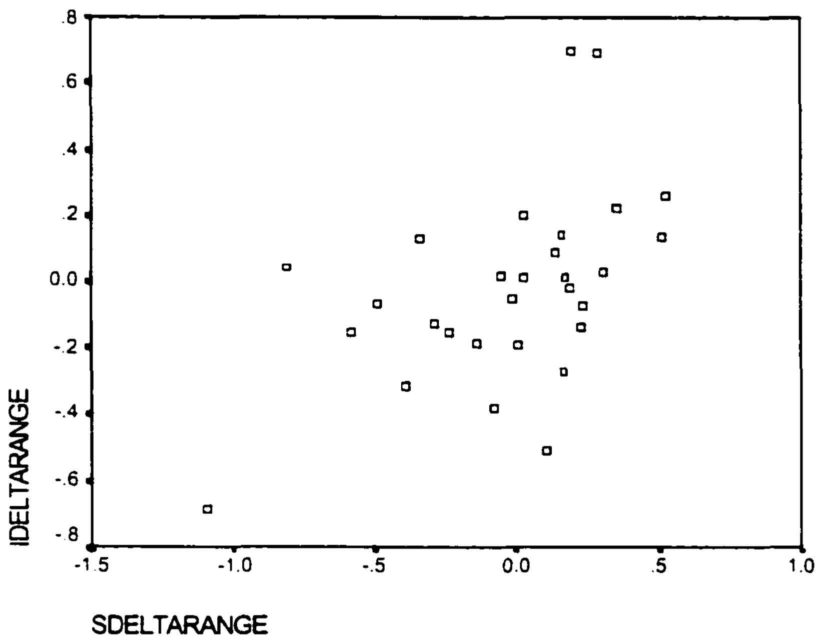


Figure 24. S10 with P, deltas, SDs.



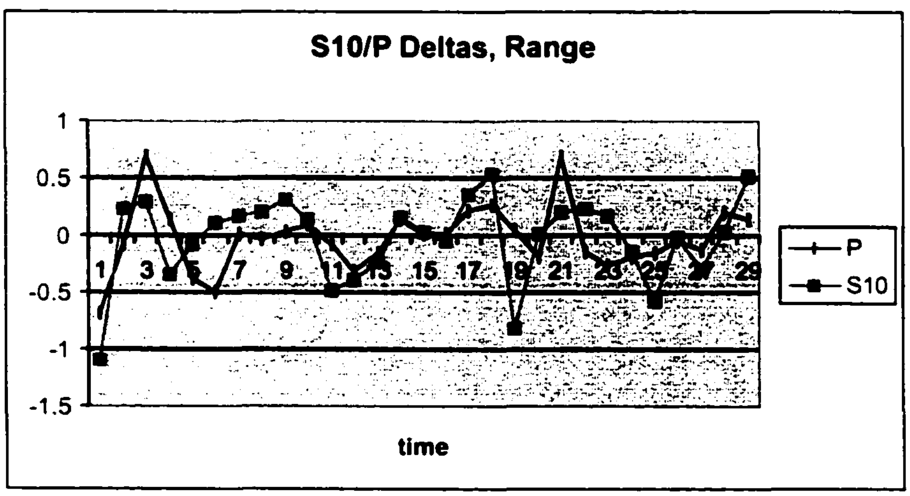
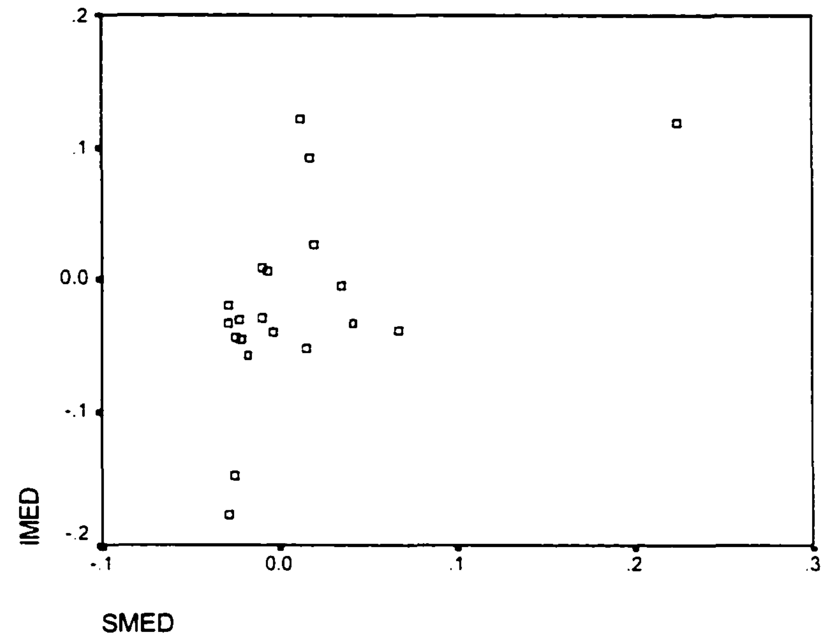


Figure 25. S10 with P, deltas, ranges.



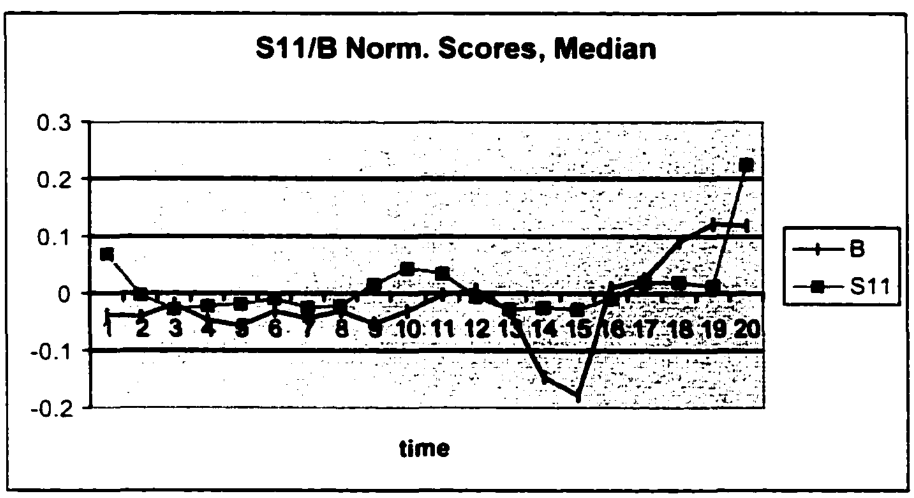
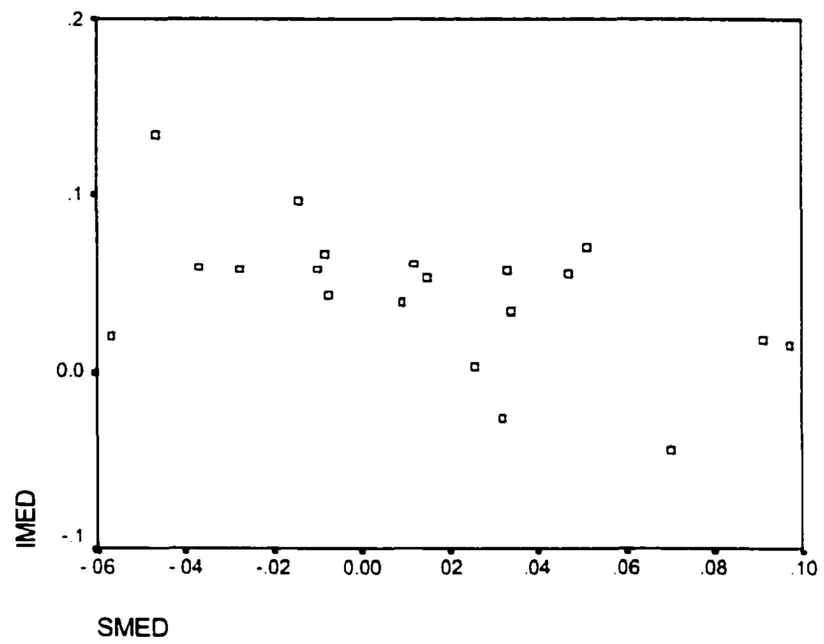


Figure 26. S11 with B, normalized scores, medians.



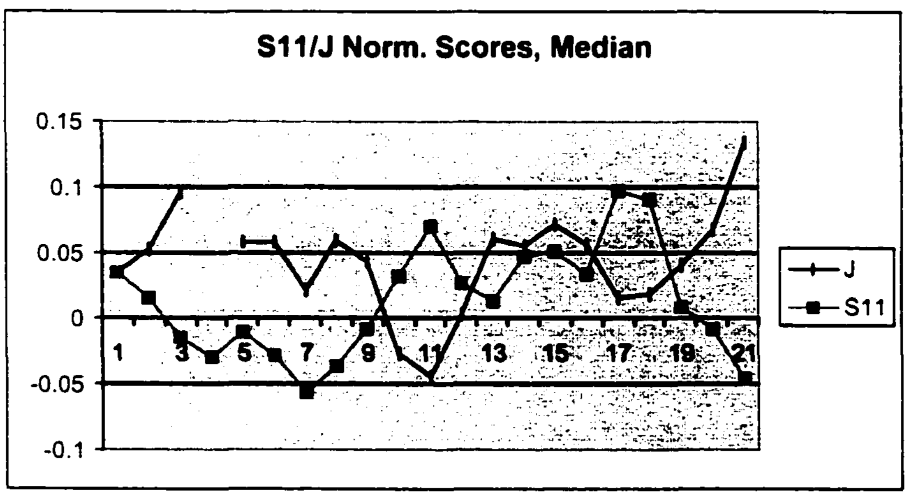
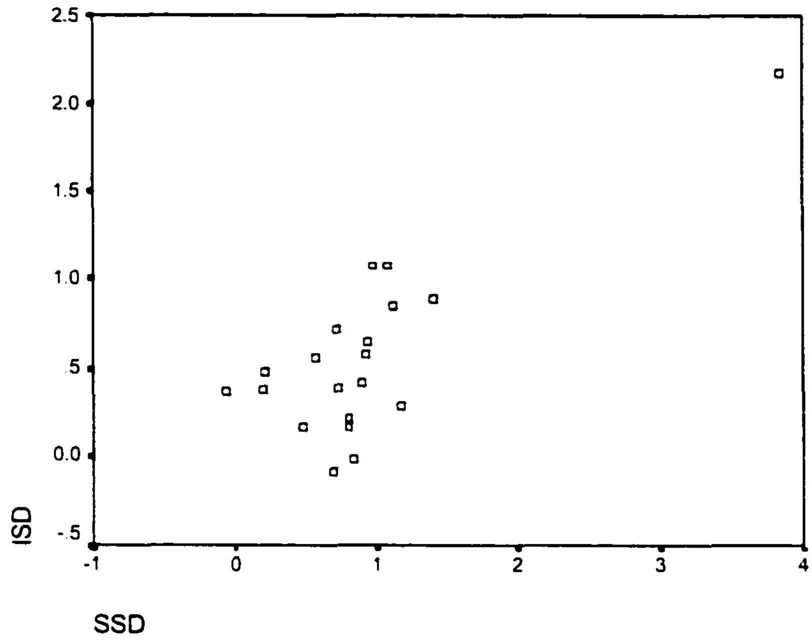


Figure 27. S11 with J, normalized scores, medians.



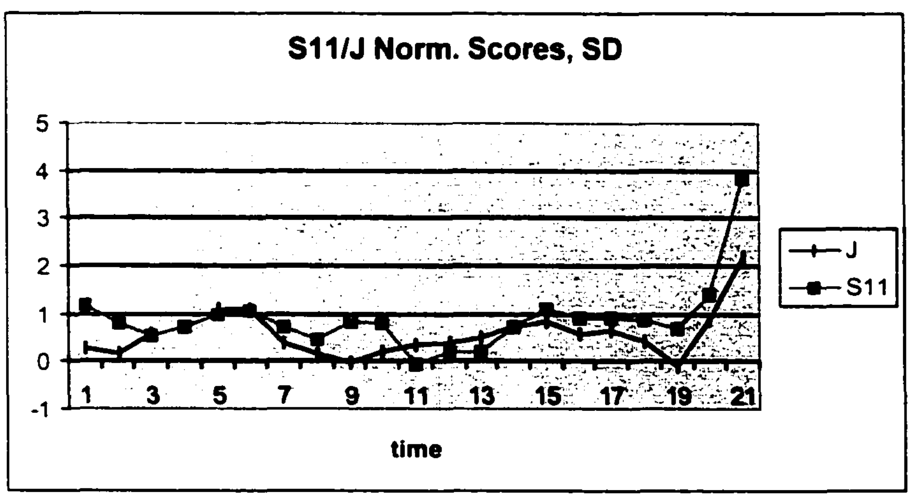
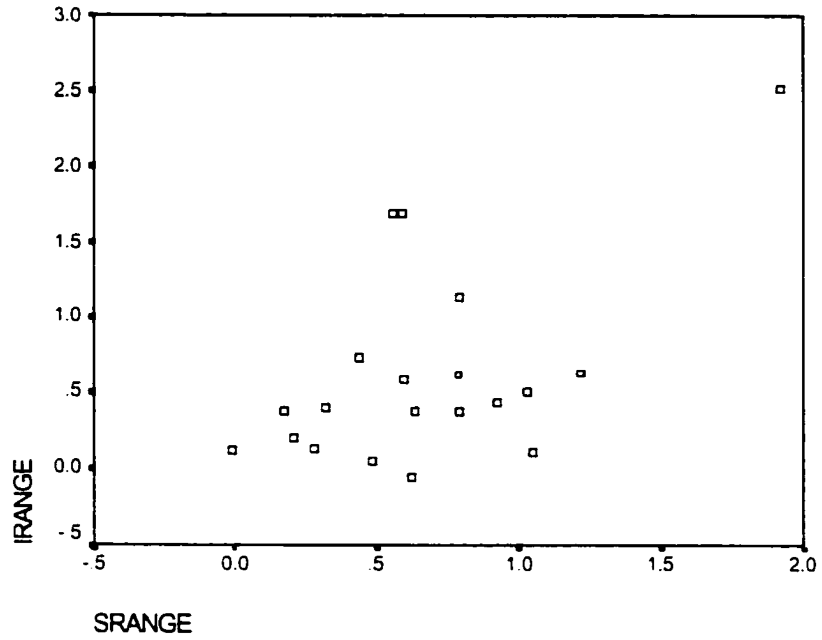


Figure 28. S11 with J, normalized scores, SDs.



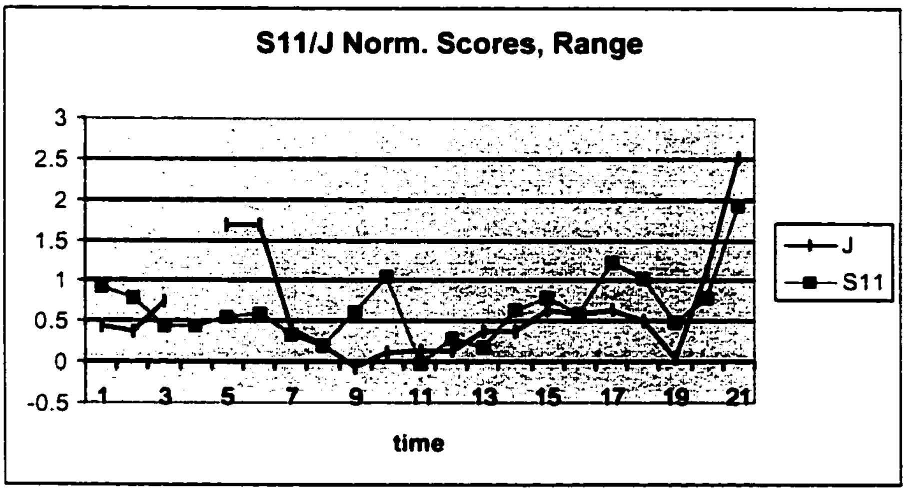
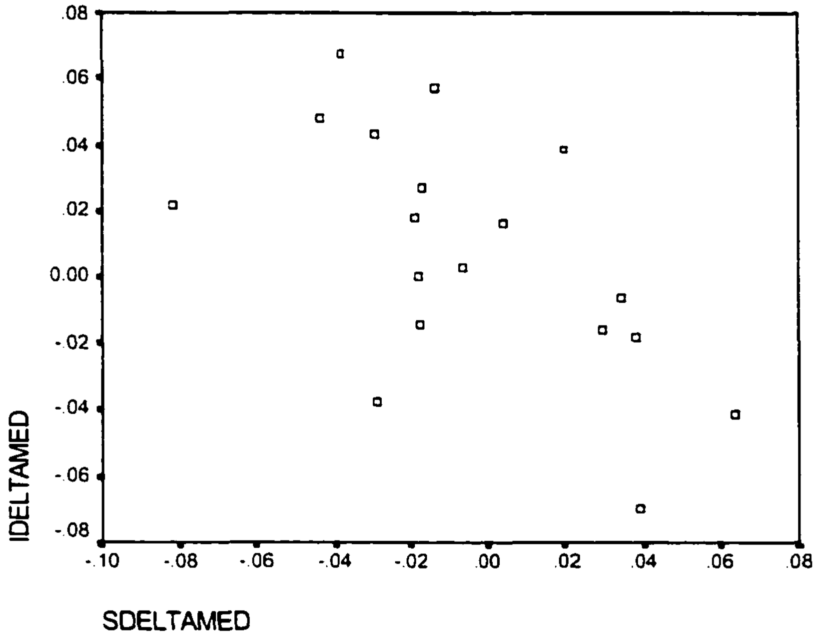


Figure 29. S11 with J, normalized scores, ranges.



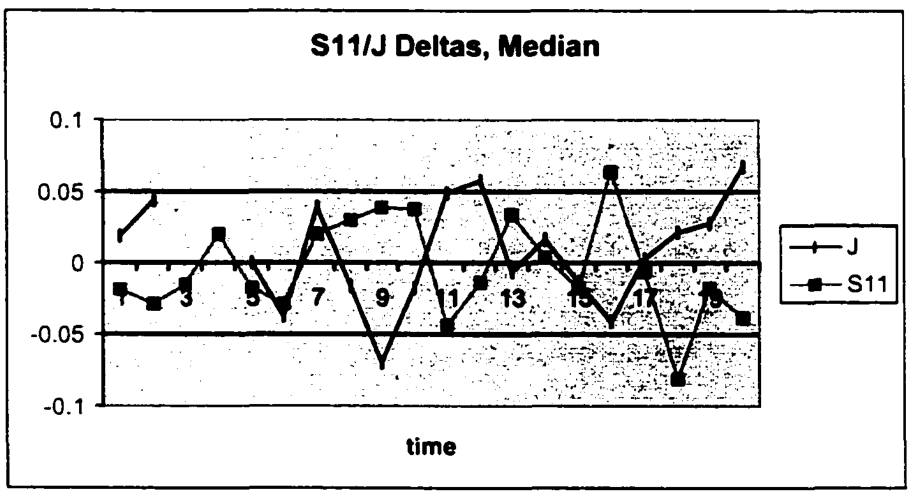
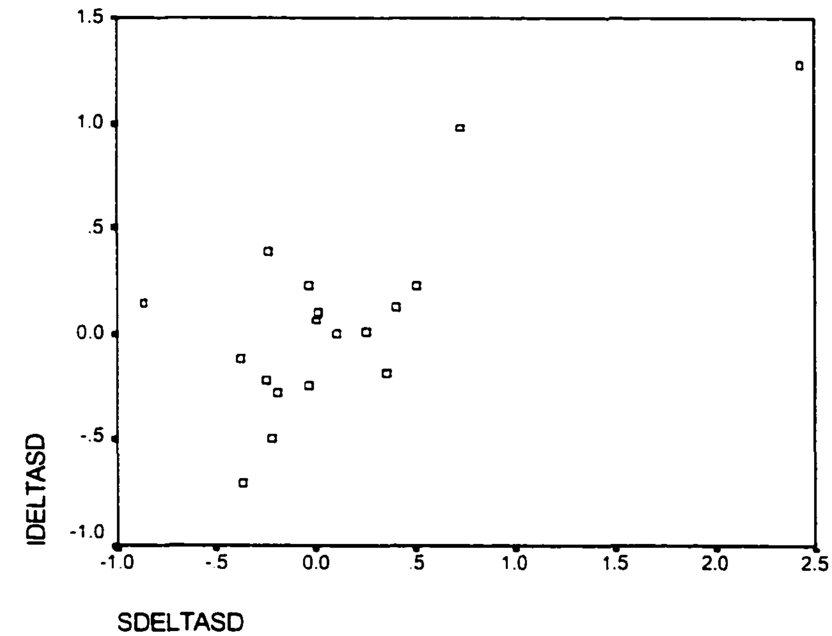


Figure 30. S11 with J, deltas, medians.



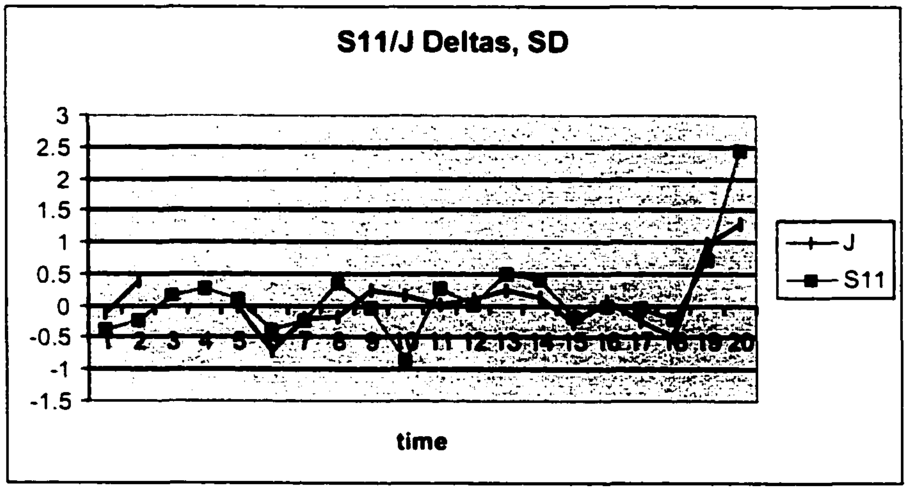
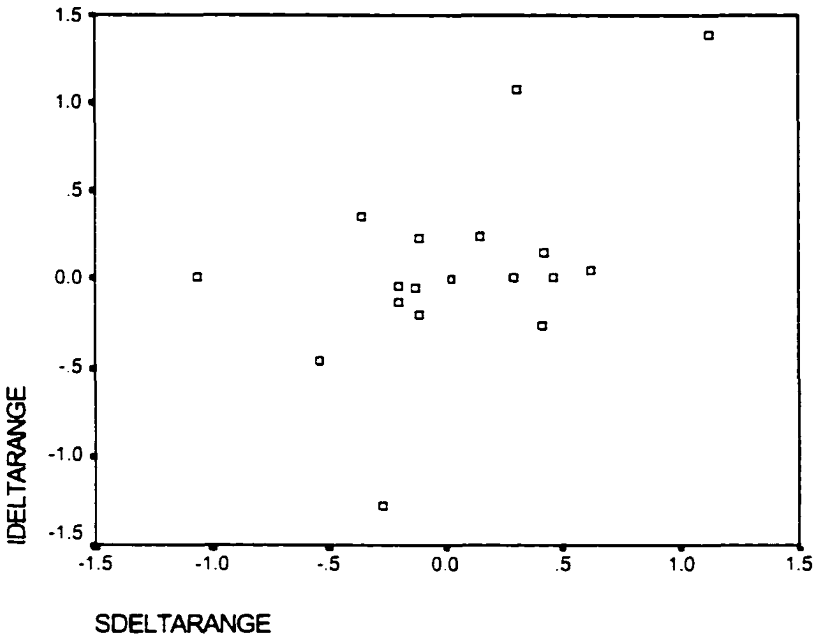


Figure 31. S11 with J, deltas, SDs.



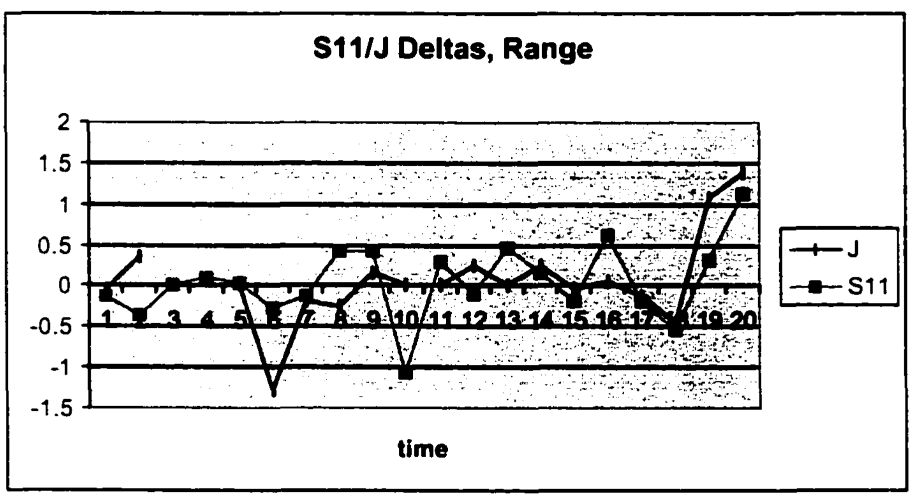
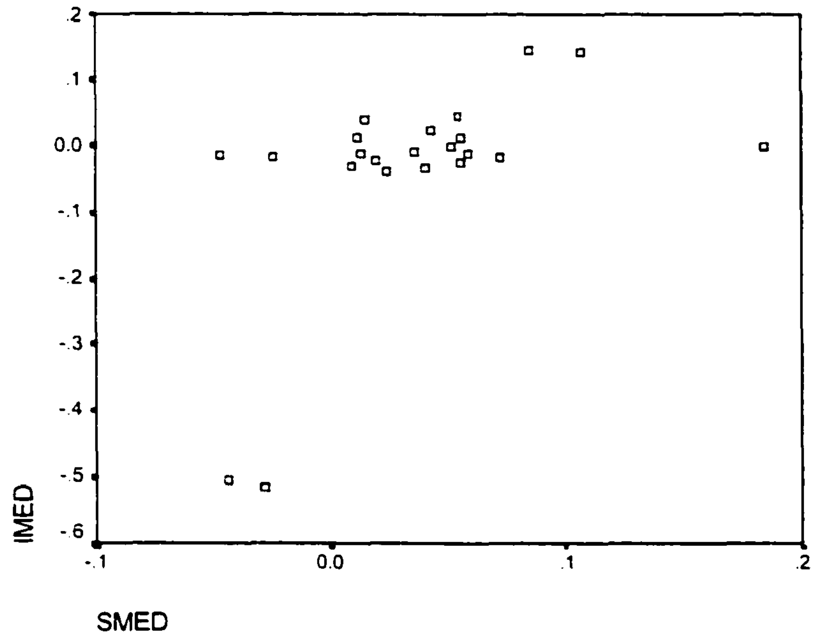


Figure 32. S11 with J, deltas, ranges.



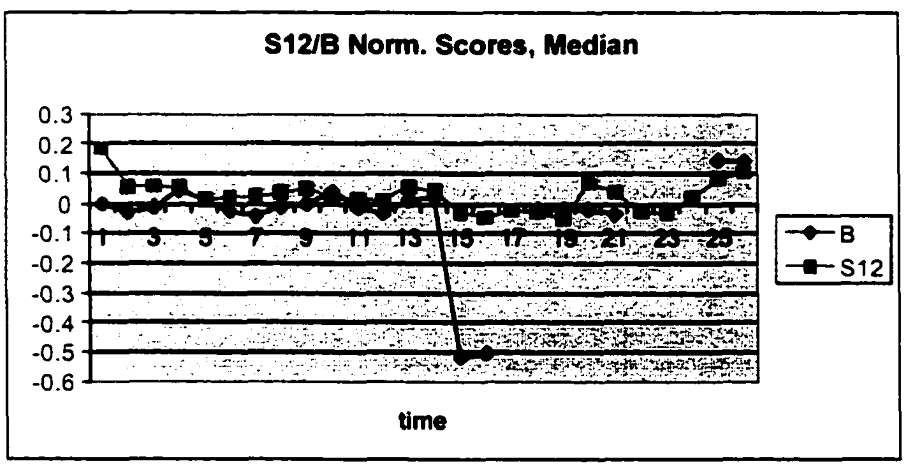


Figure 33. S12 with B, normalized scores, medians.

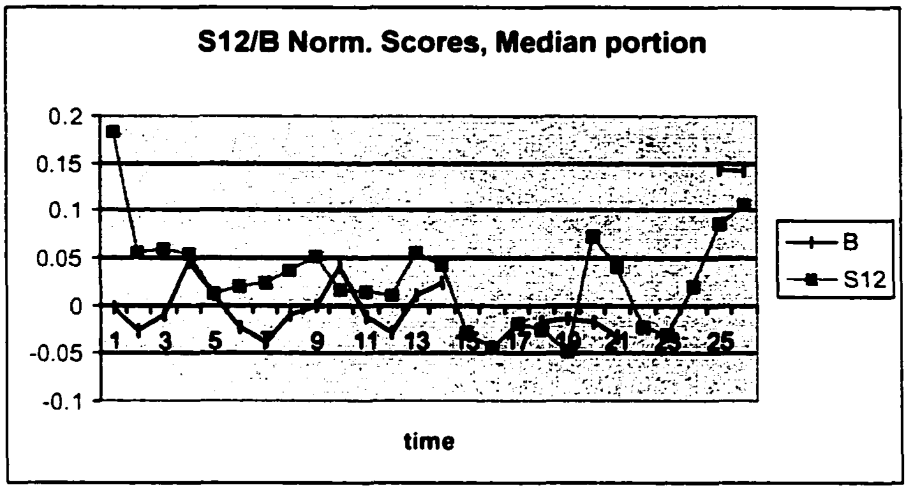
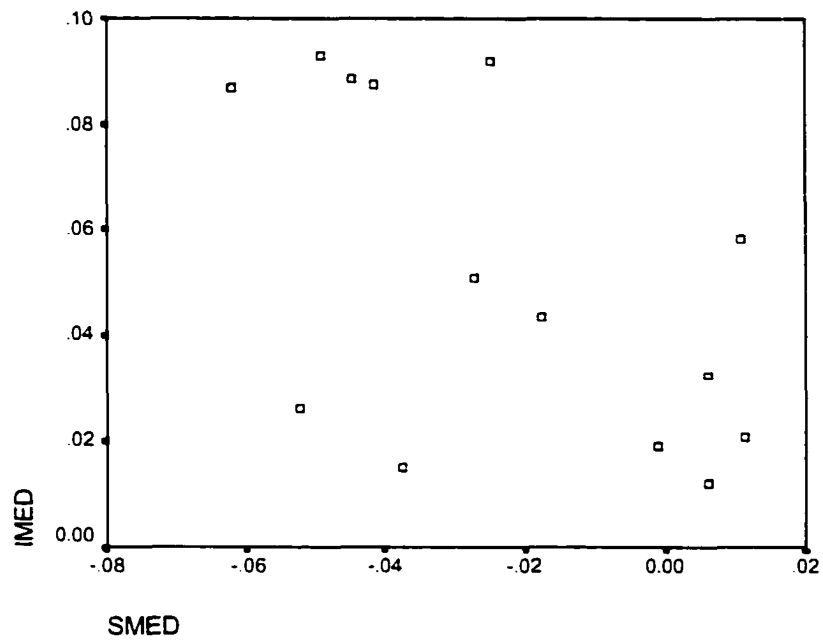


Figure 33b. S12 with B, normalized scores, medians, without possible outlier (still correlates).



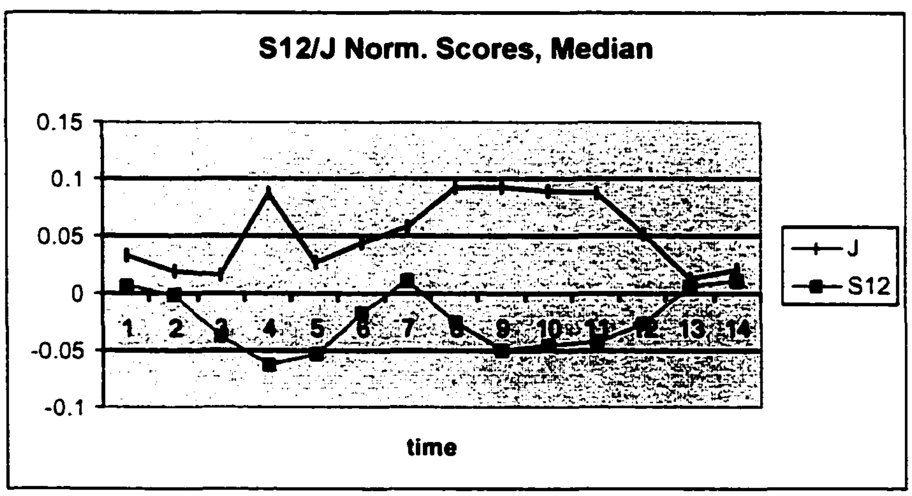
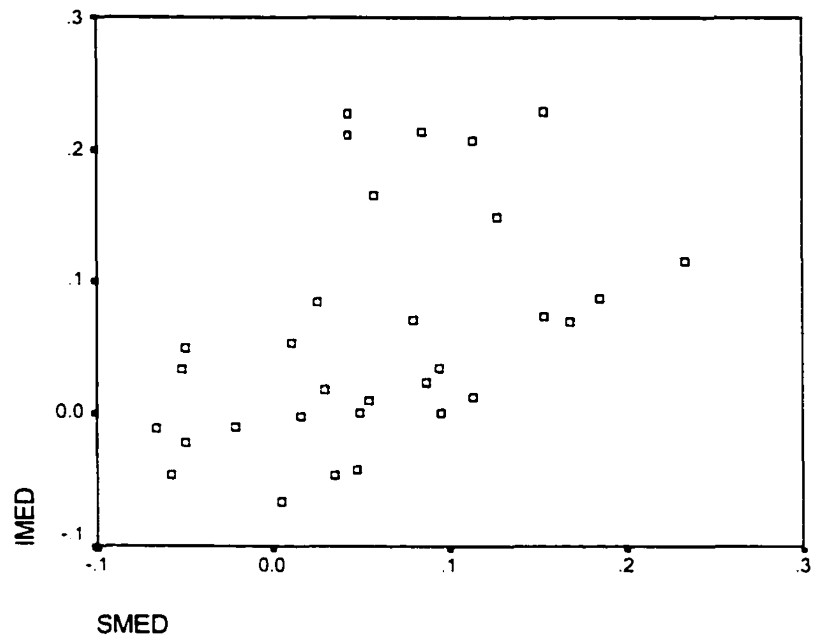


Figure 34. S12 with J, normalized scores, medians.



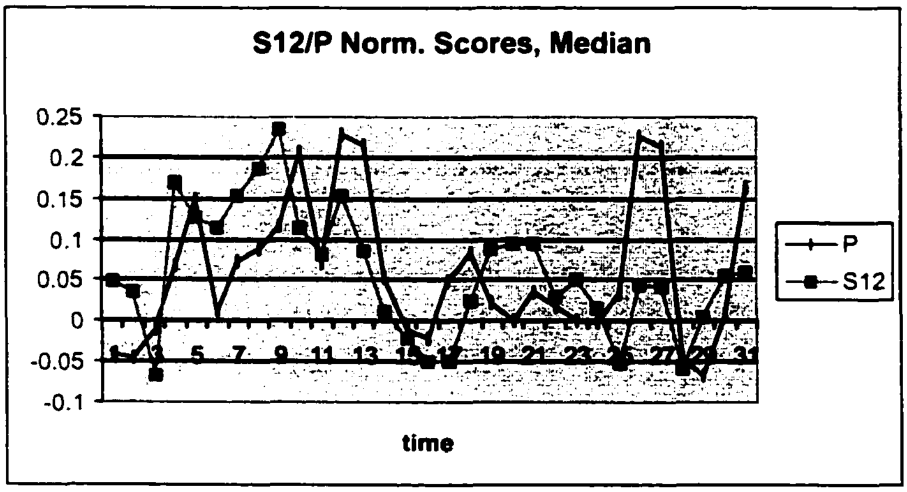
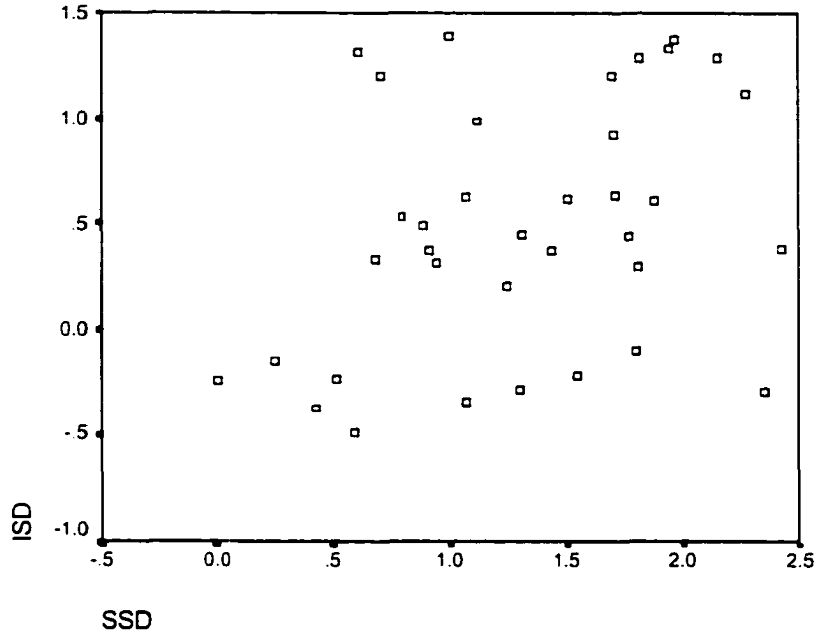


Figure 35. S12 with P, normalized scores, medians.



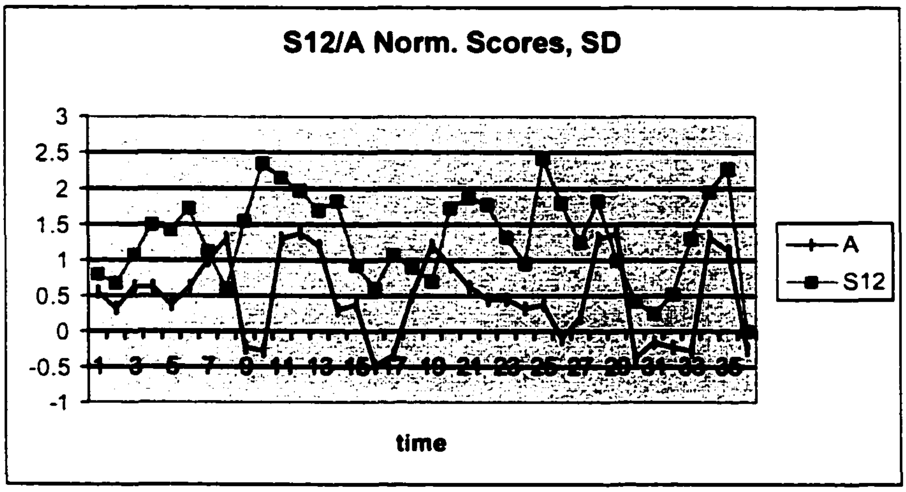
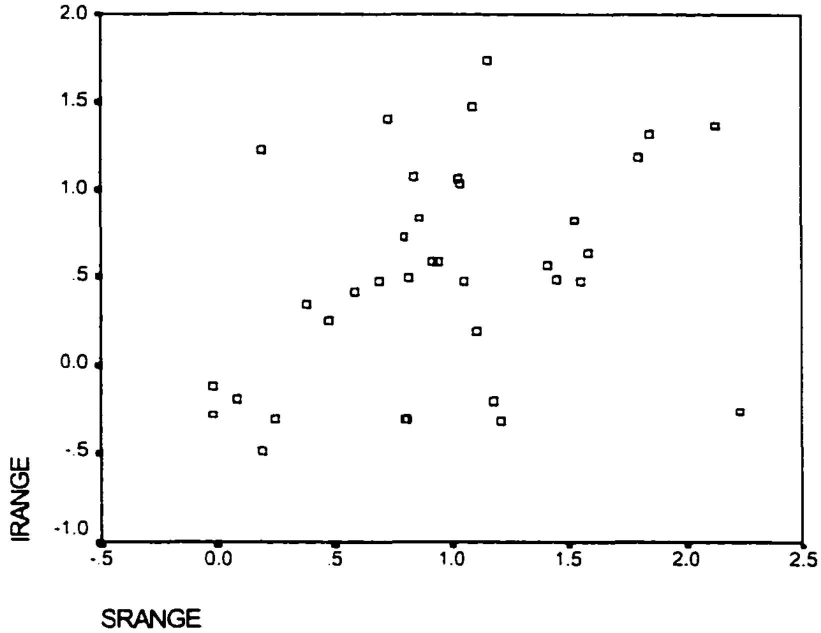


Figure 36. S12 with A, normalized scores, SDs.



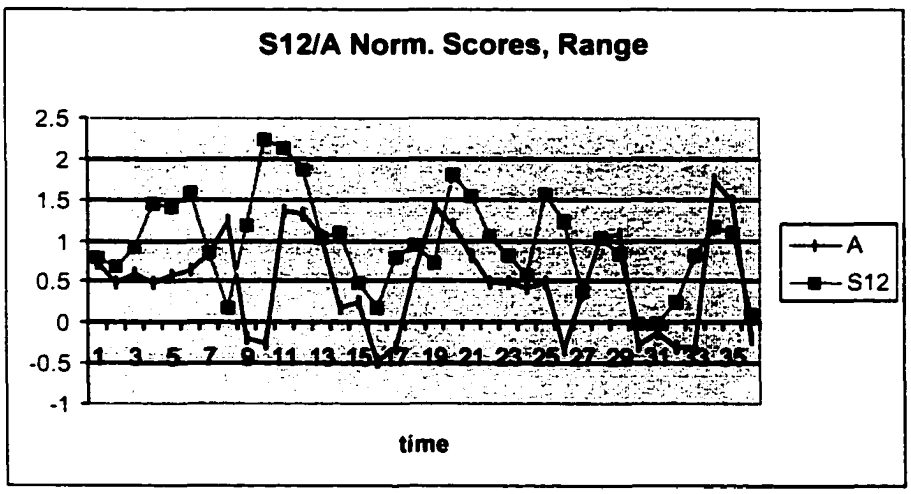
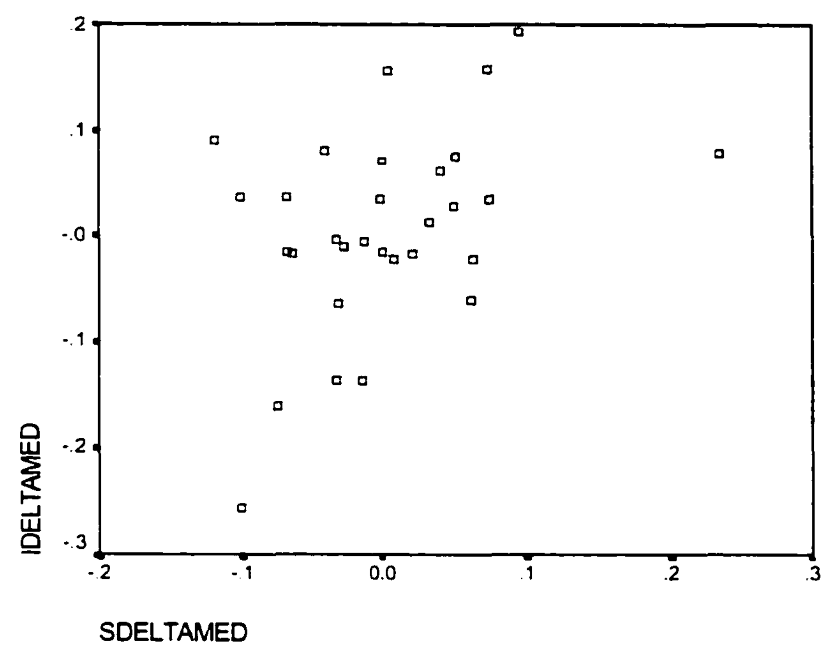


Figure 37. S12 with A, normalized scores, ranges.



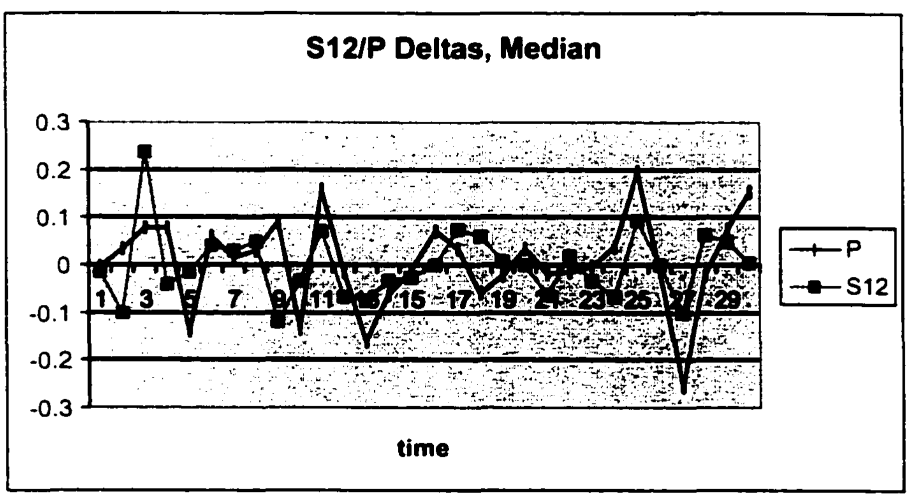
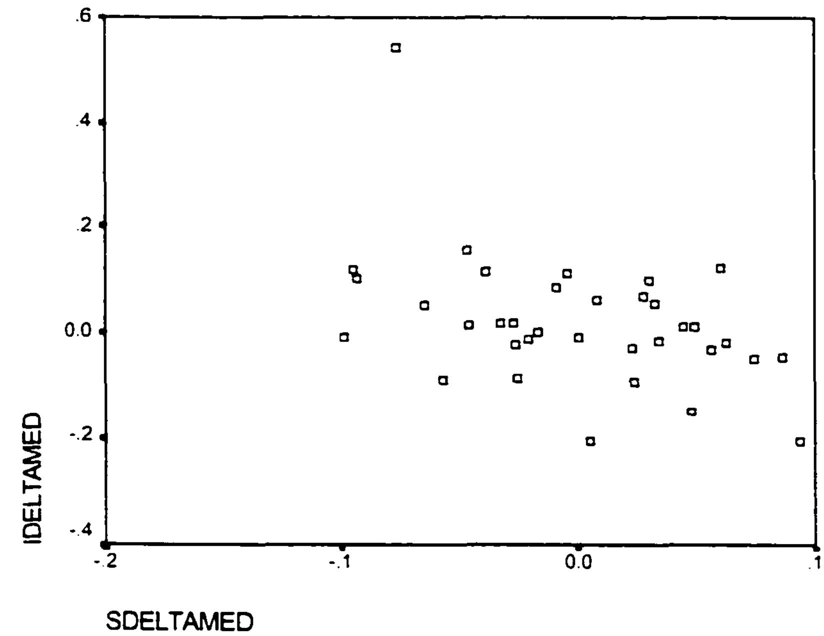


Figure 38. S12 with P, deltas, medians.



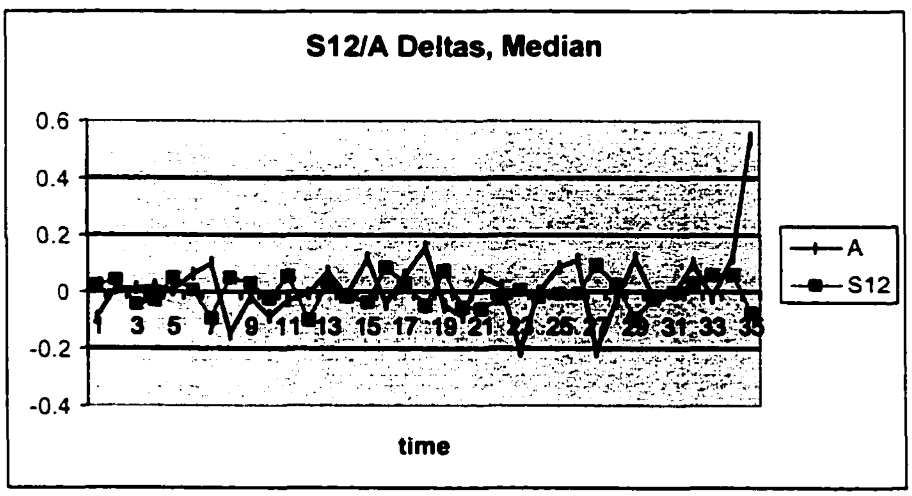
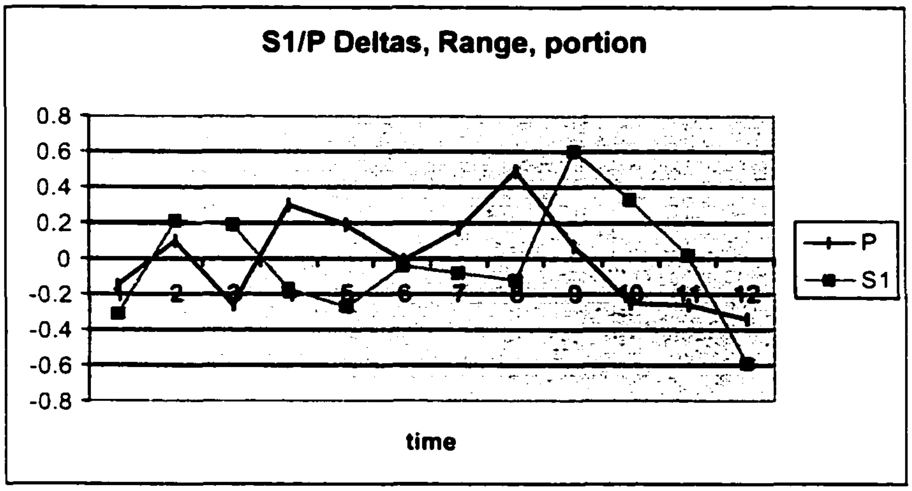


Figure 39. S12 with A, deltas, medians.

Appendix 6: Suspicious Cases without Outliers, Graphs



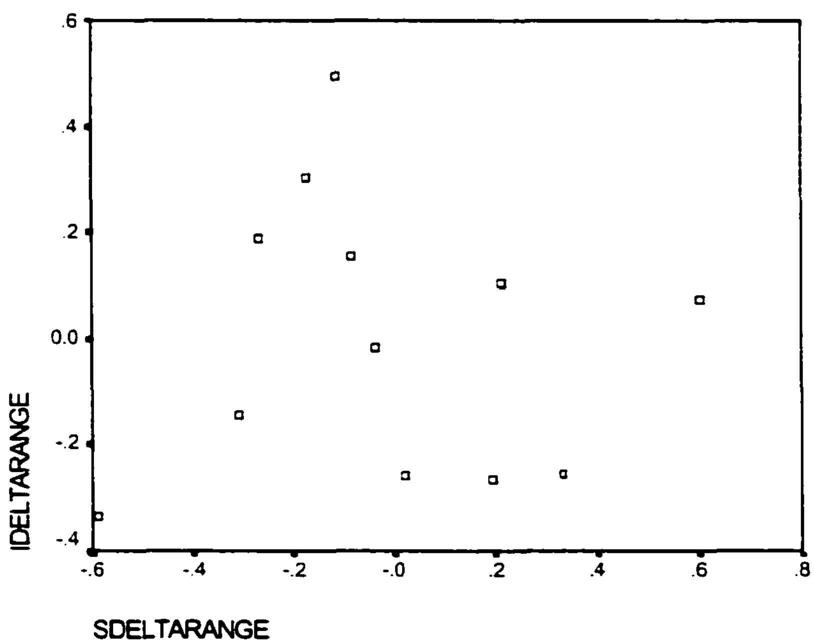
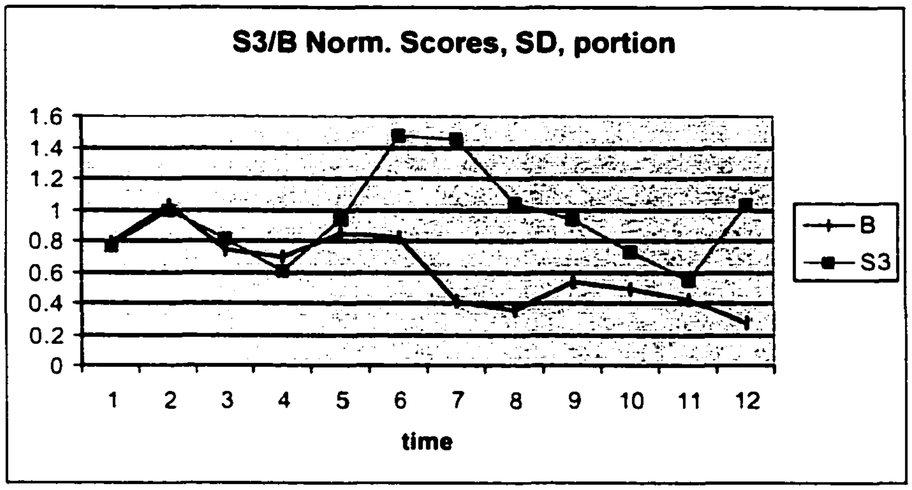


Figure 1. S1 with P, deltas, ranges, without outliers.



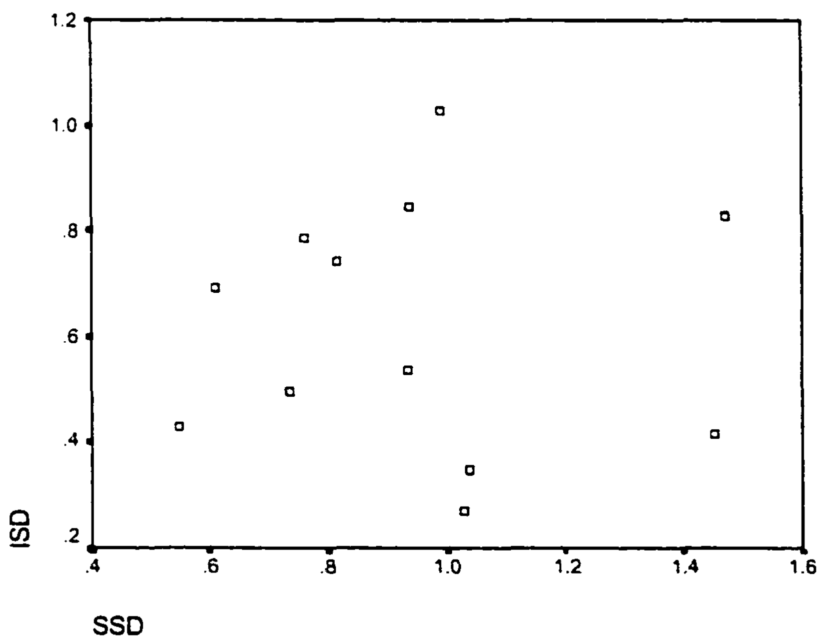
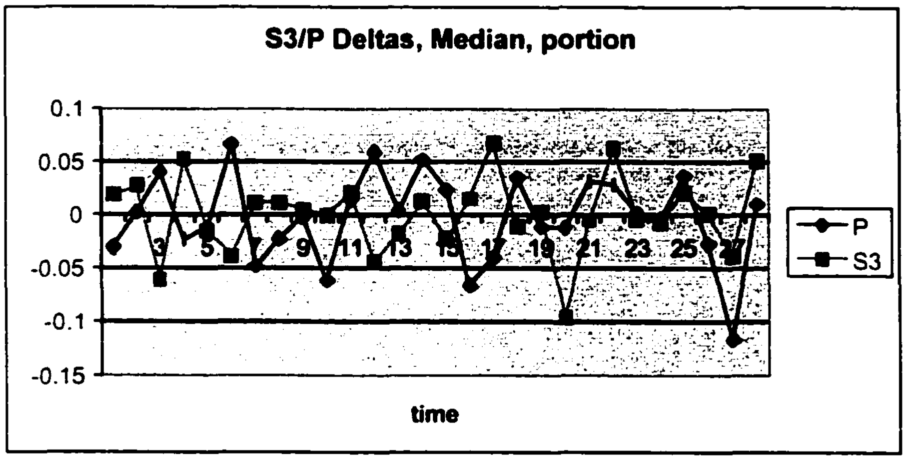


Figure 2. S3 with B, normalized scores, SDs, without outlier.



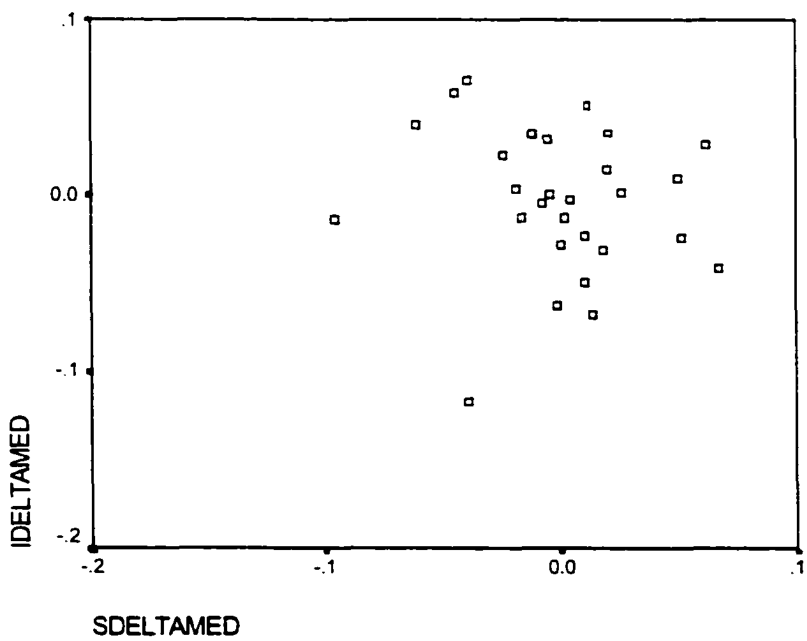
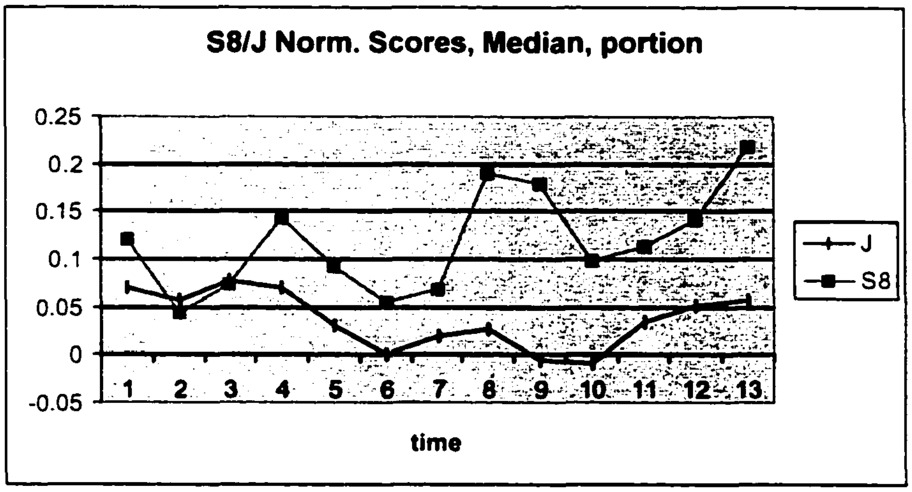


Figure 3. S3 with P, deltas, medians, without outlier.



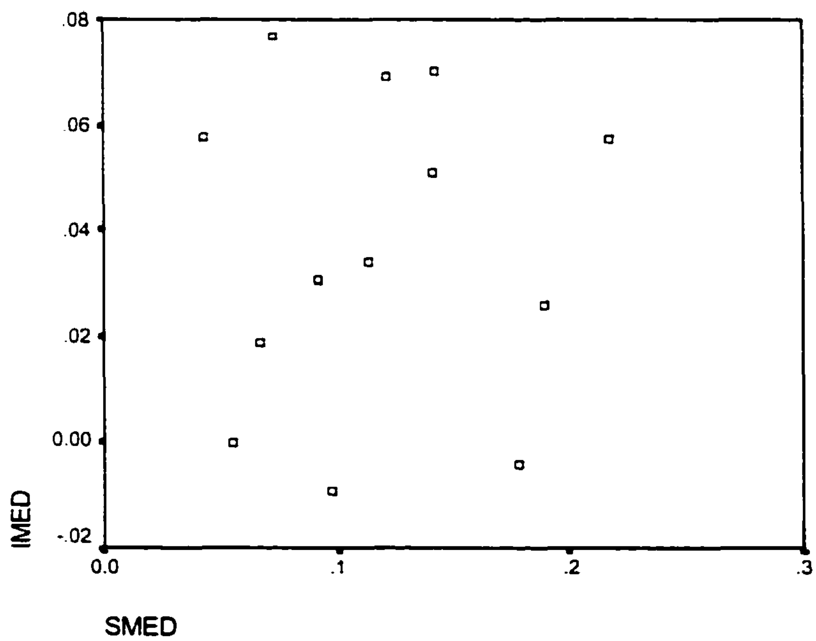
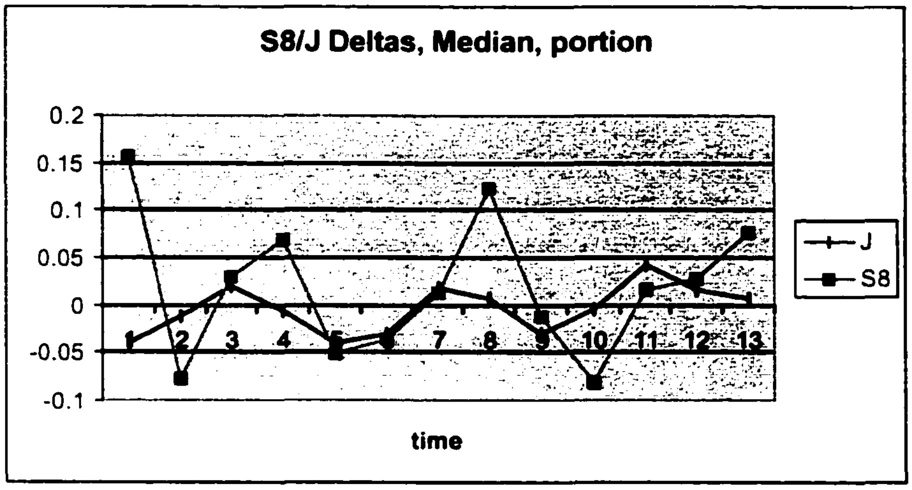


Figure 4. S8 with J, normalized scores, medians, without outliers.



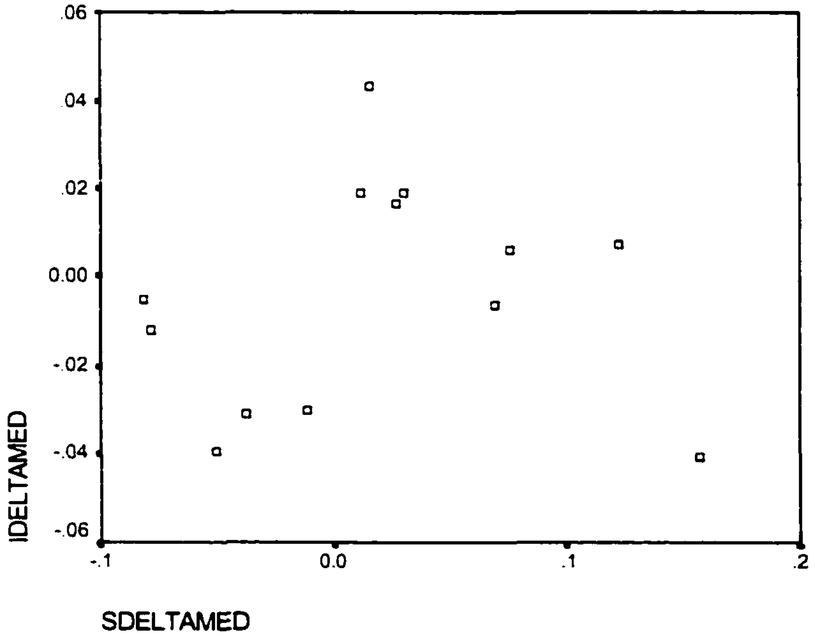
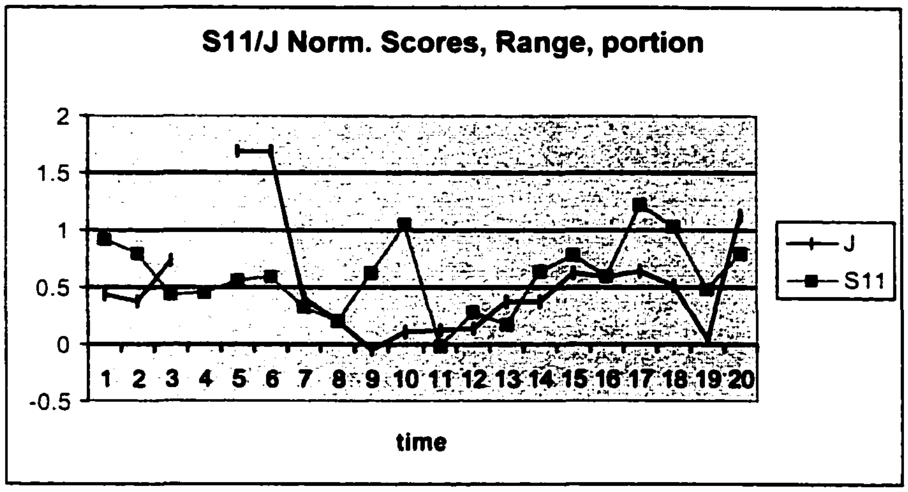


Figure 5. S8 with J, deltas, medians, without outlier.



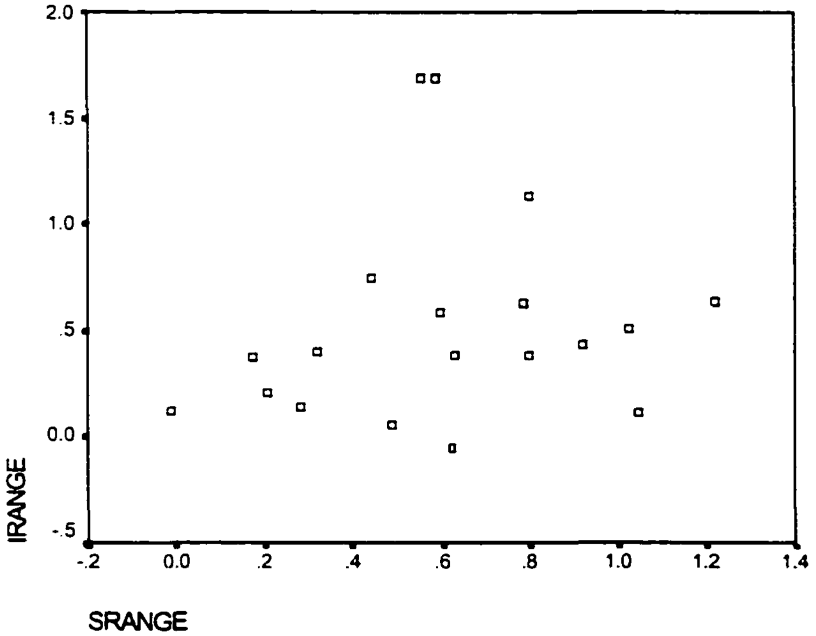


Figure 6. S11 with J, normalized scores, ranges, without outlier.

Appendix 7: Summary of Research on Female Breathy Voice Quality

My original dissertation topic was an exploration of whether or not voice quality, specifically breathiness, is manipulated (on whatever level of consciousness) as a social cue to stance, attitude, personality, etc., and via these cues acts as yet another tool or signaling device in the construction of femininity. The hypothesis was that breathiness as well as pitch might vary according to the gender and status of the interlocutor and/or possibly according to the personality or gender identity of–and thus the communicative strategies utilized by–different women. Technical difficulties that arose in carrying out this analysis prevented me from carrying it out. But, breathiness may well be important in the phonetic construction of femininity. Therefore, I will summarize both the literature and the issues that led to my experimental design for the breathiness study and discuss briefly the difficulties that eventually led me to abandon it.

Several studies have found that women's voice are breathier than men's voices. In Södersten and Lindestad's (1990) study, speech pathologists judged the degree of glottal closure via fiberscopy and listened to recordings to judge breathiness (rated on a scale from one to seven) in Swedish speakers sustaining the vowel /i:/. Incomplete glottal closure was significantly higher for females than for males. It was found in 94.5% of the females' samples (69 of 73) vs. in 37.5% (27 of 72) for the males. For both men and women, degree of incomplete closure increased with decreasing loudness, but it was only significant for the women. Perceptually, there was a significantly higher degree of breathiness in female voices than in male voices; 68 out of 69 females' vowels were rated above 0 (at least slightly breathy), while for the males, 19 out of 74 were rated as having no breathiness (0 on the scale) and only 3 were higher than 1 on the scale. No female closed her glottis completely at habitual pitch and loudness (p. 609).

Inverse filtering, a technique that involves filtering out the resonances of the vocal tract, enables researchers to examine the glottal source wave more directly. Price (1989) used this technique and found characteristics of females' glottal sources that are compatible with the above findings that women are breathier. She found that women have shorter closed quotients (percentage of time the vocal cords are closed during the glottal cycle): CQ was 40% for men, but just 16% for women. The acoustic ramification of this shorter CQ is that source-filter interactions will occur over a greater proportion of the glottal cycle for women, as the chambers (subglottal and supraglottal) are coupled when the glottis is open and not when it is closed. This interaction may be responsible for wider formant bandwidths in women. She also found that women have steeper spectral slopes in the middle of vowels; curves fall of at 1 dB per 48 Hz for men vs. 1 dB per 37 Hz for women (averages). These findings all corroborate women being more breathy, on average. Results from other studies, Hanson (1997), Holmberg et al. (1995), Karlsson (1986), and Klatt and Klatt (1990), are consistent with these findings.

While women are breathier, men tend to have more creak (Henton and Bladon, 1988, cited in Henton, 1999). Klatt and Klatt (1990) also mention that [a] breathy-laryngealized termination is characteristic of many male speakers and may be a social marker of maleness (p. 821). Within Glasgow English, Stuart-Smith (1999) found that men had more nasalization and creaky voice, and women had more whispery voice in read speech.

Just as with pitch, undoubtedly both physiological and social factors are involved in the differential usage of these voice qualities by men and women, and the questions regarding the location of the line between them and how the two kinds of factors interact remain open. A few researchers have suggested physiological explanations for women being breathier than men (e.g., Forchhammer, 1974, cited in Södersten and Lindestad, 1990, to explain the posterior glottal chink found via fiberscopy in normal female speakers, claims that some females may have functionally insufficient interarytenoid muscles) but these theories remain largely unsubstantiated. A possible motivation for socially manipulating breathiness is that it leads to a more tone-like sound (i.e., with quieter harmonics above the fundamental frequency) which could evoke a sympathetic response from hearers for ethological reasons (Ohala, 1994). Henton and Bladon (1985) suggest another ethological motivation; they suggest breathiness could be part of a courtship display ritual that imitates the voice quality associated with arousal. R. Lakoff suggests it is used to signal intimacy (personal communication). Regardless of the nature of the physiological or social motivations for breathiness, isolating social manipulations of it by testing whether there are intra-subject differences that depend on the social context (such as who the interlocutor is and what the topic of conversation is) would begin to shed light on and tease apart these factors. Thus, it was my hope to find a way to rigorously measure breathiness in recorded conversations such as the ones utilized in my pitch study.

Researchers have made significant progress in finding acoustic correlates of breathiness. Several sources describe spectral correlates of breathiness that are based on the finding that breathy-voiced vowels have steeper spectral slopes. Hanson (1997) describes several measurements that reflect spectral slope: Larger differences between the first two harmonics, between the first harmonic and the first formant, and between the first harmonic and the third formant all correlate with a more open glottis and thus a breathier voice. Her methodology involves using low-mid vowels for measurements in order to ensure separation of the first harmonic and the first formant and an equation that corrects for the effects of the vocal tract transfer function on the harmonics. She also cites larger bandwidths in lower formants and smaller harmonics-to-noise ratios of the third formant as indicators of breathy voice. Hanson's methodology and findings are promising, but her subjects were using careful speech in a laboratory setting.

Hillenbrand, Cleveland, and Erickson (1994) explored the use of a different kind of spectral measure that indicates the periodicity of a signal, Cepstral Peak Prominence (CPP), and found that it predicted perceived breathiness fairly accurately. This measure involves the cepstrum, which is the spectrum of a spectrum, of a signal in relation to a regression line drawn through the cepstral magnitudes as a normalization. It is useful in that breathy signals are less periodic and thus will have a smaller CPP than modally-voiced signals. CPP and other spectral measures were used together by Blankenship (1997) who studied both contrastive and non-contrastive breathiness in vowels in several languages. Smith (1999) also used a variation of CPP, comparing the peak to the average amplitude of quefrencies above a certain point, in her study of the prosody of utterance-final vowels in French. Her study points to another, discursive kind of linguistic manipulation of voice quality.

In addition to using spectral correlates to identify breathy voice, I hoped to find acoustic correlates of breathiness in the temporal domain, i.e., measures based on the waveform (amplitude over time), so that the two kinds of evidence could be used to corroborate each other. Ladefoged and Antoñanzas-Barroso (1985) developed a time-domain method that measures breathiness by indicating the noisiness of the signal. Their method explores the amount randomness in the waveform by making autocorrelation measurements within individual cycles of the waveform and then averaging the best-fit measurements. This method is suggestive, but the perception side of the study found that the breathiness index it creates is less important perceptually than spectral tilt measures. However, the listeners were English speakers and the stimuli were created from recordings of speakers of !Xóõ, a Khoisan language spoken in the Kalahari desert.

Other time-domain methods were considered with the idea of calibrating them with the Electroglottograph. For instance, if the opening of the vocal folds in each cycle creates a smaller second peak in the waveform, the Open Quotient could be measured indirectly by filtering the waveform around the third formant and finding this peak (John Ohala, personal communication, Hanson and Chuang, 1999). Another idea was to adapt Ladefoged and Antoñanzas-Barroso's within-cycle autocorrelation method to measure the F3 frequency's noisiness, assuming that the autocorrelation will be less precise during the open phase of a cycle due to interaction with the subglottal cavity (damping) (Ohala, personal communication). Both of these time-domain measures focus on the periodicity in the waveform *and* the shape of the glottal pulse and thus would triangulate well with several of the spectral measures discussed above, which also reflect the shape of the glottal waveform, as opposed to Ladefoged and Antoñanzas-Barroso's idea which measures the noise component of the waveform.

Many of the above studies' analyses were based on isolated vowels or syllables. Natural speech complicates their application. It would be extremely difficult to control all of the segmental, pitch, and other effects that could confound the measurement of breathiness. Hillenbrand and Houde (1996) did apply the techniques described in Hillenbrand et al. (1994) to continuous speech with some success, but their speakers all read the same sentence. In addition to the controls discussed in Hanson (1997), such as using low-mid vowels to minimize F1's effect on the lowest harmonics and controlling the segmental environment of the samples, I considered controlling for pitch either by creating a pitch library for each subject and comparing samples at the same frequencies, or by manipulating the pitch to create monotonic samples without any aperiodic material via the PSOLA (Pitch Synchronous Overlap and Add) application in Praat (software developed by P. Boersma and D. Weenink). I would then make a long-term-average spectrum of concatenations of the samples thusly created. The hope would be that segmental effects would be neutralized if the samples were representative and long enough.

Other researchers have wrestled with these issues, as well. Epstein (2001) concatenated duplicates of pitch periods to stabilize pitch and vowel quality before making spectral measurements. Her methodology could be used on real speech, especially if it could be automated. Another new development is the software developed by Smith and Robinson (2000, 2001) that can segment a waveform according to its glottal phases by identifying spectral changepoints during pitch periods. This software could be used to triangulate with other methods.

I did not feel that I was in a position to evaluate whether or not results obtained from such methods would be valid. I was concerned that the margin of error in results from these fledgling methodologies could be larger than the margin of difference that I might be trying to uncover in my data. It may be the case that more engineering needs to be done before linguists can use these methods, or scholars with both computational/signal processing and linguistic knowledge could either carry out the research or verify the validity of the measures.

It is not just these acoustic measures that need to be validated. There is a more theoretical problem to be surmounted before breathiness can be adequately understood and measured, pointed out by Kreiman and Gerratt (2000). They suggest that it is circular to define a vocal quality such as breathiness in terms of acoustic or physiological correlates and then validate such measures by their correlation with perceptual measures. Perceptual measures must be independently validated, which is no small task.

Clarifying the nature and causes of breathiness in female speakers is important for social and perceptual reasons as breathy voices, like high pitched ones, tend to be evaluated ambivalently as feminine yet incompetent (cf. Jacobi, 1996, for negative portrayals of breathy voices). Furthermore, this knowledge is also important for the endeavor to synthesize natural sounding female speech. It may well be that voice quality differences between men and women hold the key to natural-sounding synthesis.