

Chapter 3

Methods for exploring the social consequences of listening

Because listening, like speaking, is a ubiquitous aspect of language use, its study has been approached for a wide range of reasons and with a wide range of methodologies. Chapter 1 discussed some of the theoretical approaches that have been taken towards the role of the listener in linguistic variation and change, with particular emphasis on socially meaningful variation. This chapter turns from the theoretical background of variation and its significance to the methodological history of the techniques employed in the study.

The core approach is a minimally paired version of the Matched Guise Technique (MGT), a tool developed by Lambert and his colleagues in the 1960's (Lambert *et al.* 1960; Anisfeld *et al.* 1962; Lambert *et al.* 1965; Lambert 1967; Tucker and Lambert 1969). Chapter 2 showed a very basic example of this technique, used in the pilot work. Although it provided interesting and useful information, this reduced version falls far short of the social complexity needed to fully explore the role of (ING) in constructing meaning. A much richer and more nuanced set of data must be gathered. Each of the individual techniques employed in the main study has a history within the existing MGT literature: using digitally manipulated speech, creating stimuli from spontaneous as opposed to read speech, conducting open-ended group interviews both for data gathering and as the basis for a survey approach, using a Web-based survey

to widen the geographic scope of the study. None of these techniques can yet be described as routine (although for some of them it is to be hoped that they soon will be), however all have been employed before. The methodological significance of this study is that it combines these techniques and that it does so in pursuit of a novel goal: to uncover the social meaning of a single linguistic variable from the perceptual point of view in a nuanced as well as rigorous manner.

This chapter describes the methodology employed in the current study and positions it within the literature. I first review the history of the MGT and its relatives, developed within the field generally known as “language attitudes”. Although I provide a broad overview of the topics, languages and language varieties which have been explored using these techniques, I take particular note of techniques and innovations which researchers have used to augment the validity and depth of these approaches over the years. After describing the methodological history on which this research is based, I explain in detail the methods used in the present study.

3.1 Previous research on language attitudes

The Matched Guise Technique is typically employed in order to learn about the attitudes that speakers have towards language, both their own and that of others. Many worlds of complexity hidden behind those two words, “language” and “attitude”. As sociolinguists we are intimately familiar with the multitude of levels on which language may vary. When talking about differences in relation to language, we may be referring to variation occurring at level of the sound, morpheme, word, phrase, discourse or in paralinguistic cues. Working along another dimension, differences may arise with respect to specific individual variables, sets of variables correlated into styles, whole language varieties or completely separate languages. Speaker/listeners have attitudes towards and reactions to variation along all of these dimensions and more. This makes the field of studying attitudes towards language an incredibly rich and interesting one, but also means that the questions to be asked and ways finding the answers will vary tremendously depending on which specific aspect of language is under investigation.

The concept of attitude in the psychology literature is likewise a flexible and multilayered one. One commonly quoted definition is that attitudes are “a disposition to react favorably or unfavorably to a class of objects” (Sarnoff 1970:279), however recent work in the larger framework of attitudes has suggested that this is overly simple, in that attitudes may “not always be simply positive or negative, but may subsume both positivity and negativity” (Haddock and Maio 2004b:1). There is also increasing evidence that not only do people differ in which attitudes they hold, but even in how their attitudes are structured (Haddock and Huskinson 2004).

Attitudes are generally held to have three components: **affect** (how we feel about French, or about a Southern accent), **cognition** (the beliefs we hold about them) and **behavior** (how we speak and react to others’ speech). As a result, attitudes are diverse and sometimes contradictory, shifting based on how they’re measured, the context of measurement and a range of interpersonal factors. Although measures of these three components (affect, cognition and behavior) do correlate under many circumstances, under others they don’t (Haddock and Huskinson 2004), making it difficult to be sure exactly what the stable, underlying character of attitude is, or if one exists.

As a construct, attitudes have been primarily explored within the field of social psychology, which devotes itself to a deeper understanding of the mental processes involved in interpersonal behavior of various sorts. Studies in language attitudes have been carried out by social psychologists working in this context who have a particular interest in language or aspects of language. They have also been conducted by sociolinguists whose primary focus is the structure of language and its relationship to social constructs and processes (see Thomas (2002) for a review of sociolinguistic uses of perception studies). Although these two threads of research have slightly different perspectives, they share a great deal of common interest. Through the years, there have been repeated efforts to bring these two fields into closer alignment (Giles 1992; Milroy and Preston 1999; Edwards 1999). Social psychologists are best able to contribute a detailed understanding of the psychological processes involved in the evaluative reactions and other social behaviors. Conversely, sociolinguists are uniquely qualified to explore the evaluative aspects of specific forms of variation and

other socially meaningful aspects of language. Edwards (1999) points out that we “would benefit, therefore, from efforts to bridge the work of psychology and linguistics in this regard; the effect would be to refine and particularize our knowledge of how specific aspects of speech elicit specific types of evaluative reactions” (p. 105).

Because of the complexities of both of these definitional elements of the field, theoretical and methodological difficulties are intricately tied together in the study of language attitudes (Agheyisi and Fishman 1970; Lee 1971; Forgas 1983; Giles and Bourhis 1976; Ryan *et al.* 1982). In the following sections, I describe the main thrust of the literature on language attitudes, focusing particularly on the use of the MGT and its methodological history and innovations. After that, I discuss the particular tools in the literature which I have adapted for the current study and describe their importance, before moving on to describe my own methodology in detail.

Methods for exploring language attitudes

The study of language attitudes or listener perceptions encapsulates a range of questions and methods. The most direct method of investigating attitudes is to ask speakers to tell their opinions through surveys, interviews and questionnaires. This approach may be used on its own and has been in such diverse populations as Arabic university students (Zughoul and Taminian 1984; Al-Haq 2000), participants in a multilingual school in Spain (Fitch and Hopper 1983), French-speaking Walloons in Belgium (LefÈVre 1979; LefÈVre 1978), Flemings in Belgium (Geerts *et al.* 1978), French and English speakers in Canada (Bourhis 1983; ClÉment and Noels 1992; Landry and Bourhis 1997), Italian and English speakers in Canada (Bourhis and Sachdev 1984), English speakers in Detroit (Shuy and Williams 1973), Puerto Ricans in New York City (Fishman *et al.* 1971; Wolfram 1973), residents of Alsace (Vassberg 1993), Israelis and Palestinians in Israel (Kraemer 1992), Vietnamese speakers in Australia (Pittam *et al.* 1991), students in Wales (Roberts and Williams 1980) and about topics such as beliefs about language and gender in the U.S. (Kramer 1977; Kramer 1978; Scott 1980) and attitudes of U.S. teachers (Taylor 1973) and others (Hoover 1978; Speicher and McMahon 1992) towards African-American Vernacular

English and of residents of Scotland towards English and Scots (Pollner 1987), residents of Zimbabwe towards Shona and English (Mparutsa *et al.* 1992), Australian aboriginal children towards Aboriginal English and Standard Australian English (Purdie *et al.* 2002) and Australians of Anglo, Italian and Greek descent towards a variety of ethnic groups (Callan and Gallois 1983). In some cases, it is even possible to use overt questioning to investigate speaker attitudes towards individual variables, as in Boberg (1999), which explores listener beliefs about the appropriate pronunciation of the written letter (a) in foreign loan words in American English. A more detailed approach to the overt questioning method is language diaries, which Lawson and Sachdev (2004) combined with a more general questionnaire to investigate the language choices and code mixing patterns of Sylheti-Bengali students in London.

The field of perceptual dialectology is based on a particular subset of the direct question method, focusing on the individual speakers' beliefs regarding regional variation (Preston 1986; Preston 1989; Williams *et al.* 1996; Preston 1999b; Preston 1999a; Iann`Accaro and Dell'Aquila 2001; Long 2002; Preston and Robinson 2005). One of the crucial contributions of perceptual dialectology is its emphasis on the structural aspects of speaker/listener beliefs. Although the entire field of language attitudes is devoted to the study of speaker beliefs regarding language, much of that study presupposes both the linguistic and social structures to be investigated. In contrast, work in perceptual dialectology is devoted to uncovering what divisions speakers make between regions on a linguistic basis and what terms and descriptions they use in labeling the linguistic varieties they have delineated. Preston (1989) makes this point in describing how one would interpret results of a hypothetical matched guise study examining listener reactions to various U.S. regional dialects: "though it is accurate to report, in terms of production dialectology, that the response was to a New England voice, it is surely as important (perhaps more) to determine where the respondents believed the voice was from and eventually important to relate that to the group's overall taxonomy of dialect differences" (p. 3).

In pursuit of this goal, researchers in perceptual dialectology have employed a range of tasks, the most common one being the map task, in which respondents are

given a relatively blank map of their country or region and asked to draw lines indicating which areas contain people who speak the same way. Other tasks include asking respondents to rate individual areas (such as U.S. states) for speech qualities such as correctness and pleasantness, as well as the “pile-and-sort” method in which participants sort cards labelled with states to prove them into piles of similar linguistic patterns. These kinds of approaches have been used to investigate speaker beliefs about the regional divisions within the U.S. as a whole (Preston 1989; Williams *et al.* 1996), Ohio in particular (Benson 2003), Ladin-speaking southern Europe (Iannàccaro and Dell’Aquila 2001), Wales (Williams *et al.* 1996) and Great Britain more generally (Inoue 1996), among others.

Although overt questioning can produce interesting results on its own, it may also be used within larger sociolinguistic projects. For example, Labov (2001) discusses a comprehensive study of variation in Philadelphia, using interviews which culminated in explicit questions regarding speech in the area. Many studies combine overt questions with the MGT or verbal guise approaches discussed below (d’Anglejan and Tucker 1973; Williams 1973; Kerkhoff *et al.* 1988; Garrett *et al.* 2003). Ladegaard (2000) goes even further, by combining general attitude questionnaires with a verbal guise task as well as correlating these with actual linguistic behavior, investigating the attitudes and uses of teens in Denmark toward the vernacular of the area. The overt approach is primarily useful when examining attitudes towards easily conceptualized units of language, for example separate languages, language varieties or speech in specific geographic areas.

Another approach for learning about the attitudes present in an area or community is content or media analysis. This does not examine the beliefs of individual speakers, but rather looks at discourses available in the larger culture by studying demographic patterns or historical events and their impact on linguistic beliefs and behaviors, as well as the the themes that emerge from literature, media and other public sources (St Clair 1982). Work in this domain involves analyzing, for example, the deployment of accents in movies (Lippi-Green 1997). On a larger scale, researchers may explore sociohistorical and demographic trends and how they affect beliefs and attitudes regarding a particular language such as French (Bourhis 1982), Chinese (Li 2004),

varieties such as New Zealand English (Gordon 1989) or a topic such as the history of language legislation in Belgium (Geerts 1988) or the interaction of language and gender (Kramarae 1982). This approach, too, may be combined with others, for example Stevens (1983) uses general social issues to supplement his direct knowledge of attitudes towards Arabic and French gained from interviews and ethnographic observation in Tunisia.

These two methods of inquiry focus on beliefs and ideologies that speakers may articulate about language, either at the individual level or within cultural discourses. This information does not necessarily shed light on the role that linguistic traits play in day-to-day individual interactions. To address that question, we need data which reveals more about how individuals listen to each other's linguistic performances and respond to them. In the 1960's, Lambert and his colleagues developed the Matched Guise Technique to covertly elicit individuals' attitudes towards members of different ethnolinguistic groups (Lambert *et al.* 1960; Lambert *et al.* 1965; Lambert 1967; Tucker and Lambert 1969). The technique involves having a single speaker produce two (or more) recorded stimuli in different languages or varieties. For example, a bilingual speaker might record the same passage read aloud in French and English or a bidialectal speaker might read a passage aloud twice in English, once with a standard accent and once with an accent which is regionally or socially marked. Participants in the study are asked to listen to the recorded speakers and evaluate them on a range of qualities, for example, how intelligent, educated, friendly or trustworthy they sound.

Because listeners are not told that the alternate recordings have been produced by the same person, they evaluate each "guise" (language or accent performance) as an individual speaker. However, because the recordings have been produced by the same person, many of the paralinguistic cues are (hopefully) held constant, for example speech rate, pitch contours and various aspects of voice quality. Likewise, since the speaker is usually recorded reading aloud the same passage, the content of the utterance is also held constant. As a result, it is possible to (at least tentatively) assign any differences in the evaluations to ideas that the listeners have regarding the specific languages or varieties being used.

This basic innovation was remarkably brilliant and relatively straightforward and

quickly spawned an entire body of work. As researchers adopted the basic idea to a range of goals and situations, they introduced new aspects to the approach. In addition, critiques of the very real limitations of the technique (Agheyisi and Fishman 1970; Lee 1971) led to refinements and improvements. The following discussion will detail the situations in which this technique has been employed and the adaptations that have been made to it.

Contrasting languages, varieties and variants

Research using the MGT has been carried out on a range of linguistic situations including multilingual settings (Edwards 1983), for example exploring attitudes towards Hebrew and Arabic in Israel (Lambert *et al.* 1965), French and English in Canada (Lambert *et al.* 1960; Genesee and Holobow 1989), Spanish and Quechua in Peru (Wölck 1973), Tamil and Kannada in India (Sridhara 1984), Castilian and Catalan in Spain (Woolard 1984; Woolard and Gahng 1990), Swiss German and High German in Switzerland (Hogg *et al.* 1984) and English, Cantonese and code switching in Hong Kong (Gibbons 1983). In addition to competing languages, the MGT has been used to investigate attitudes towards regional or social varieties such as regional accents in England (Strongman and Woosley 1967; Giles 1971a; Giles *et al.* 1983a; Giles *et al.* 1990; Giles *et al.* 1992; Dixon *et al.* 2002), Welsh accent and RP in England (Giles 1971b; Creber and Giles 1983; Brown *et al.* 1985) and Wales (Price *et al.* 1983; Garrett *et al.* 2003), Indian accents in England (Elwell *et al.* 1984), English and Scottish varieties in Scotland (Cheyne 1970; Abrams and Hogg 1987), different regional accents in Ireland (Edwards 1977), Hawaiian Creole English and Standard American English in Hawaii (Ohama *et al.* 2000), Spanish-accented English in the U.S. (Mckirnan and Hamayan 1984), Chicano English and Standard American English in L.A. (Arthur *et al.* 1974; Bradac and Wisegarver 1984), gendered perceptions of female English speakers (Batstone and Tuomi 1981; Giles *et al.* 1980), Appalachian English in the U.S. (Luhman 1990), French Canadian accents in Canadian English (Webster and Kramer 1968), Jewish accents in Canada (Anisfeld *et al.* 1962), “broad” and “refined” Australian accents (Ball *et al.*

1984) as well as global and non-native varieties of English in Australia (Ball 1983; Callan and Gallois 1982; Seggie 1983), standard American and Chinese-accented English in the U.S. (Cargile 1997) and Japanese-accented English in the U.S. (Rubin *et al.* 1991; Cargile and Giles 1997; Cargile and Giles 1998). Researchers have also investigated reactions to class-based linguistic variation in Ireland (Edwards 1979) and French-speaking Canada (d'Anglejan and Tucker 1973) and levels of formality in Canadian French (Taylor and Clément 1974). Race has also been a significant topic, particularly in the U.S. (Fraser 1973; Johnson and Buttny 1982; Purnell *et al.* 1999; White *et al.* 1998), as has age, particularly in the U.K. (Giles *et al.* 1990; Ryan and Laurie 1990).

In early work, Giles (1973) used an interesting version of the technique to establish the basic principle of Accomodation Theory, that speakers change their performances based on their interlocutors. Rather than using deliberately designed stimuli, he conducted interviews with a series of teenagers from Bristol, arranging an interview first with himself then with a peer of the students. Playing excerpts of these interviews for listeners from both Bristol and Wales, he established that the switch in interviewers created linguistic changes in the teen subjects which were perceptible to others.

When the varieties to be studied are not usually spoken in the same area, or when the goal is to compare more than two or three varieties in a single study, it may be difficult to find speakers who can convincingly portray all of the necessary guises. To investigate such situations, researchers have turned to the “verbal guise” paradigm (Cooper 1975). This version of the matched guise technique dispenses with the controlling factor of using the same speaker in favor of more believable performances. Although it makes it more difficult to be confident that it is the varieties themselves which are causing differences in evaluation, much useful information can still be obtained in this way. This technique has been used to investigate responses to children’s speech (Sachs *et al.* 1973; Edwards 1979) and multiple levels of standard or nonstandard speech including French speakers in Canada (d'Anglejan and Tucker 1973), socioeconomic variation in the U.S. (Harms 1967; Shuy 1973) and Costa Rica (Berk-Seligson 1984), reactions to Black and White speakers in the U.S. (Hopper and Williams 1973; Fraser 1973; Rosenthal 1974; Irwin 1977;

Graff *et al.* 1986; Larimer *et al.* 1988; Baugh 1996), regional U.S. accents (Smith and Bailey 1980), nonnative accents in the U.S. (Stewart *et al.* 1985; Gill 1994), regional accents in the U.S. (Frazer 1987), Hawaiian Creole English (Slaughter 1982; Yamamoto and Hargrove 1982), Spanish accents in the U.S. (Ryan and Carranza 1975; Ryan *et al.* 1977; Zerda and Hopper 1979; Podbresky *et al.* 1990; Dailey *et al.* 2005), Asian accents in the U.S. (Podbresky *et al.* 1990), Cantonese-accented Mandarin in China (Kalmar *et al.* 1987), nonnative English and Chinese speech (White and Li 1991), regional accents of Danish (Kristiansen and Giles 1992; Ladegaard 1998a) and French (Paltridge and Giles 1984), Danish attitudes towards regional varieties of English (Ladegaard 1998b), German speakers' ability to identify world Englishes (Stephan 1997), nonnative accented English in Canada (Kalin *et al.* 1980) and the U.S. (Mulac *et al.* 1974; Wylie 1980), the accent effects of moving in adult life (Munro *et al.* 1999), local and global accents in New Zealand (Huygens and Vaughan 1983; Bayard 1991; Wilson and Bayard 1992; Bayard 1999) and Australia (Gallois and Callan 1981; Callan *et al.* 1983; Gallois and Callan 1985) and attitudes towards adults with a child-like voices cross-culturally (Montepare and Zebrowitz-Mcarthur 1987). Verbal guise techniques have also been used to document one of the basic premises of descriptive linguistics, that differences between standard and nonstandard speech are socially defined, by showing that listeners unfamiliar with a given language are unable to detect aesthetic differences between varieties (Trudgill and Giles 1976).

Early work within the verbal guise paradigm was inspired by a desire to investigate the extent to which listeners are able to accurately deduce personality characteristics from speech (Pear 1931). In one sense this research was a failure, as it revealed little in the way of consistent connections between speech and actual personality of the speaker. It did however, uncover a surprising degree of agreement across listeners as to the perceived personality traits signaled by specific qualities. Some work investigating the connection between paralinguistic cues and detailed personality factors and moods has continued to this day (Kramer 1963; Scherer 1974a; Scherer 1974b; Bezooijen *et al.* 1983; Feldstein and Sloan 1984; Brown and Bradshaw 1985).

When using different speakers it can be difficult to establish which aspects of the speech trigger which aspects of the evaluation. Brennan (1977) suggested that this

issue may be addressed by developing scales of accentedness for Spanish-accented English (also reported in Ryan (1973)). By correlating specific linguistic traits with psychophysical scaling responses, Brennan attempted to develop consistent linguistic metrics to evaluate the degree of accent present in a given speech sample, a measure which could then be correlated with other evaluative responses (Brennan *et al.* 1975; Brennan and Brennan 1981). While it is not clear that this particular approach is the right one, it makes an important point that there is likely to be more stability in examining specific attributes over a larger number of speakers than in relying on the theoretical similarities across guises of one or two multilingual speakers.

Another approach is to use written materials, in order to perfectly control the linguistic features being manipulated. While this is not particularly effective for phonetic or phonological variables, it has the advantage of eliminating such factors from the performance, as well as being easy to arrange. Written materials have been used to examine factors such as lexical diversity (Bradac and Giles 1988; Ruva and Bryant 1998), the use of quotatives (Buchstaller 2003; Buchstaller In Press; Buchstaller To Appear), effects of gendered speech styles (Kramer 1974; Erickson *et al.* 1978; Mulac *et al.* 1985; Siegler and Siegler 1976), patronizing speech towards older adults (Ryan and Cole 1990; Ryan *et al.* 1991; Giles *et al.* 1993), powerful/powerless speech (Nigro *et al.* 1989; Smith *et al.* 1998; Adkins and Brashers 1995) and various forms of hate speech (Leets and Giles 1997; Leets 2001). Some work combines written and spoken materials in order to provide a “control” for noncontent cues (Trees and Manusov 1998).

When it is not possible to manipulate the linguistic performance of a single speaker, it is often possible to investigate the interactions between linguistic cues and those from other modalities, be they visual cues (Elwell *et al.* 1984) or explicit information such as the race (Dixon *et al.* 2002) or age (Ruva and Bryant 1998) of the speaker. Rather than having alternate guises, entirely new personae may be created. This approach has been most thoroughly explored in the domain of education. Williams (1973) used videos of children combined with the same audio track to investigate the influence of visual cues of race on teacher evaluations. Conversely, Rubin and Smith (1990) showed that visual cues marking a lecturer as Asian increased undergraduates’ perceptions of the speaker’s accent and actually lowered

their comprehension of the material. Seligman *et al.* (1972) looked at teacher ratings of a combination of pictures, samples of read speech and examples of student work. They found that teacher ratings were not only influenced by each of these factors independently, but that they interacted in important ways as well.

The MGT has also been extended to look at more specific speech variables, beginning with Addington (1968), who looked at an astounding seven speech qualities, three levels of speech rate and three levels of pitch variation, all crossed with each other for both male and female speakers. Other such studies include examinations of speech rate (Scherer *et al.* 1973; Smith *et al.* 1975; Apple *et al.* 1979; Street and Brady 1982; Jr. *et al.* 1983; Brown *et al.* 1985; Giles *et al.* 1990; Ray *et al.* 1991; Giles *et al.* 1992; Ray and Zahn 1999) and pitch variation (Apple *et al.* 1979; Ray *et al.* 1991; Ray and Zahn 1999). Linguistic investigations have more often looked at the effect of a range of smaller variables combining into a more general phenomenon such as powerful/powerless language (Lind and O'Barr 1979; Hosman 1989; Gibbons *et al.* 1991; Hosman and Siltanen 1994; Adkins and Brashers 1995; Schmidt and Brigham 1996; Hosman 1997; Smith *et al.* 1998; Holtgrades and Laskey 1999; Hosman *et al.* 2002; Parton *et al.* 2002), confidence (Scherer *et al.* 1973), dynamic vs. conversational modes of speech (Schweitzer 1970; Pearce and Conklin 1971; Pearce 1971; Pearce and Brommel 1972), disfluencies (Miller and Hewgill 1964; Sereno and Hawkins 1967), language intensity (McCroskey and Mehrley 1969; Bradac *et al.* 1979; Bradac *et al.* 1980; Aune and Kikuchi 1993), lexical diversity (Bradac *et al.* 1977; Bradac *et al.* 1979; Bradac and Giles 1988; Levin *et al.* 1994), verbal immediacy, that is, how closely speakers link themselves to the topic of the message (Bradac *et al.* 1979; Bradac *et al.* 1980), patronizing or over-accommodative speech towards elders (Ryan and Cole 1990; Ryan *et al.* 1991; Edwards and Noller 1993; Giles *et al.* 1993; Smith *et al.* 1998; La Tourette and Meeks 2000; Ryan *et al.* 2000), degree of accentedness (Rubin and Smith 1990; Cargile and Giles 1998), impact of timing on interpersonal attraction (Siegman 1979), managerial style (Rubin *et al.* 1991) and factors influencing politeness including explicitness, dominance and form of argument (Clément and Noels 1992), as well as positive or negative "tone of voice" (Laplante and Ambady 2003) and nonverbal cues (Trees and Manusov 1998).

A smaller body of research explores which cues listeners use in identifying language varieties (Clopper 2000; Flanigan and Norris 2000; Clopper and Pisoni 2001; Clopper and Pisoni 2003). Many of these employ digital manipulations in order to dampen information from certain sources (for example, segmental information) in order to investigate whether listeners are still able to identify a particular variety on the basis of the remaining information (speech rate and prosody). This approach has investigated how listeners identify regional dialects of Dutch (Bezooijen and Gooskens 1999), the impact of different modes of speech on persuasive arguments (Pearce and Conklin 1971; Pearce 1971), race and sex identifications in English (Lass *et al.* 1979; Lass *et al.* 1980; Foreman 2000) or general social responses (Bezooijen 1988). A cautionary note on approaches of this sort is given by Bezooijen and Boves (1986), who point out that the elimination of vocal qualities by these techniques may not be exactly as theorized. Other work uses these approaches to learn more about how listeners process speech of different kinds, for example demonstrating that listeners exposed to a wider range of regional varieties learn better how to categorize them (Clopper and Pisoni 2004a; Clopper and Pisoni 2004b).

As advances in digital technology make it easier and less expensive to create plausible sounding synthetically manipulated tokens, sociolinguists have become increasingly interested in creating studies which explore specific variables via the MGT. Early work was only possible for paralinguistic cues such as speech rate and pitch (Apple *et al.* 1979), but in recent years this ability has been extended. Graff *et al.* (1986) took recordings from a Black speaker in Philadelphia and used digital manipulations to create versions containing regional vowel markers specific to White speech in the area. Fridland *et al.* (2004) used monosyllabic tokens to create synthetically manipulated guises with different vowel formants and used these to investigate listeners' abilities to recognize certain variants as more Southern. Podbresky *et al.* (1990) took an ingenious approach to compare standard American English speakers with nonnative speakers with audible Spanish and Asian accents. Because these varieties would have been prohibitively difficult to obtain believable matched guise examples, they adopted a verbal guise technique, but used electronic manipulation to align the speech rates and intensity levels of each recording, so as to minimize the impact of

these paralinguistic cues on the results.

Although the MGT was developed specifically to allay concerns about the role of idiosyncratic speaker differences in listener reactions, very little work has been done on the ways in which those idiosyncrasies influence the topics being investigated. Arthur *et al.* (1974) represents an excellent exception to this rule, examining the ways in which Chicano guises of different speakers were rated differently. In particular, they noted that their raters (Anglo undergraduates in Southern California) showed a wider variation of ratings between speakers when responding to the standard guises than to the Chicano guises. Rather than a straightforward distinction between negative and positive reactions, Arthur and his colleagues documented a more subtle form of discrimination. The listeners responded in more complex ways to the standard, which was the closest to their own variety and compressed the range of their responses in response to the unfamiliar and potentially stereotyped Chicano variety.

In addition to factors such as region or ethnicity, researchers have also investigated the influence of gender, both of the speaker and the listener. Although there is a significant body of work on the topic, it is not always included in general discussions of the field. Many studies examine gender as part of a larger project, for example Addington (1968), mentioned above and Edwards (1979), who found that teachers in Ireland rating speech samples of nine and ten year olds rated “nondisadvantaged” girls more favorably than their male counterparts, while no such gender difference was seen among the “disadvantaged” children. Further, the male teachers rated the disadvantaged children more favorably than their female colleagues (Edwards 1979:39).

Other studies position gender as central to the research question. Sachs *et al.* (1973) showed the degree to which linguistic differences between men and women are attributable to social processes rather than biological ones by establishing that adults can distinguish between the voices of boys and girls pre-puberty. Mulac *et al.* (1985) established that both beliefs about male and female speech and actual gendered linguistic traits influence listener ratings independently. A number of papers on language and gender, inspired by Lakoff (1973), sought to test her claims regarding the attributes and perceptual consequences of so-called “women’s speech” (Siegler

and Siegler 1976; Erickson *et al.* 1978; Newcombe and Arnkof 1979). Other work explored speakers' stereotypes concerning male and female speech and the ease with which they are identified (Kramer 1974; Kramer 1977; Kramer 1978). Batstone and Tuomi (1981) took a curious approach, carrying out a standard verbal guise study on female voices, collecting the scales to be used by asking participants about qualities relevant to sexiness in young women.

Although many of the studies discussed above incorporate gender as a minor variable, two stand out as unusually rich discussions. Elyan *et al.* (1978) found that female RP guises were rated as more feminine than their corresponding Lancashire guises, but were also judged to be more masculine in particular ways, for example lacking in warmth and higher in adventurousness and independence. Giles *et al.* (1980) explored perceptual correlates of a feminist ideological stance in female speakers. They found that in spontaneous speech (on topics purportedly unrelated to feminism), feminist speakers were perceived as more feminist, more intelligent and more sincere and were rated as less frivolous, superficial and standard-speaking. However, when limited to reading a passage aloud, many of these assessments were reversed, with the feminist speakers rated as less fluent, less standard and less intelligent.

These creative approaches to presenting stimuli to participants in order to elicit their evaluations is made necessary by the wide range of linguistic behaviors researchers have sought to examine. This diversity likewise necessitates diversity in the evaluative tasks requested of listeners, with respect to both topic and technique.

Status and solidarity

Many studies have focused on how language attitudes contribute to larger patterns of intergroup relations. Different social settings have been shown to lead to different patterns and language attitudes. In some, distinct linguistic groups maintain mutually hostile or suspicious relationships, each favoring speakers of their own varieties on skills such as honesty or friendliness (Lambert *et al.* 1965). In other cases, speakers of minority varieties seem to have internalized the negative attitudes present in their culture, rating their own varieties as badly or even worse than outsiders do (Lambert

et al. 1960; Genesee. and Holobow 1989), or they may share these negative opinions but to a lesser extent (Ohama *et al.* 2000; Dailey *et al.* 2005). Alternatively, speakers of minority varieties may note the negative status of their own variety, while upgrading its speakers on affective terms such as honesty or friendliness (Wölck 1973). These patterns can get more complex, as in Luhman (1990), where Appalachian listeners downgraded their own variety in terms of status, while favoring it in terms of solidarity, but only for the male speakers. Ryan (1979) discusses a range of studies exploring positive regard for nonstandard varieties, while Ryan *et al.* (1982) discuss the possible orientations towards a nonstandard variety more formally.

Researchers have repeatedly found connections between the perceived **ethno-linguistic vitality** of a language or variety and the covert attitudes towards it (Bourhis and Sachdev 1984; Clément and Noels 1992; Giles and Johnson 1981; Ryan *et al.* 1982). Along similar lines, the perceptions of Spanish-speaking high-school students regarding the presence of Spanish in the **linguistic landscape** (Landry and Bourhis 1997) around them correlates negatively with their evaluations of Anglo speakers (Dailey *et al.* 2005).

When reasoning about the meaning of different ratings for different varieties, it is important to remember the class associations of the particular varieties and speech styles employed in the experiment. Because many regional varieties are associated with lower socioeconomic status, it is difficult to determine which associations are being reflected in attitude responses (if, indeed, these can be separated). Abrams and Hogg (1987) argued that much of the downgrading of regional varieties, particularly with respect to status, reflected class assumptions which went unexplored in the original studies. In their study, which used two varieties of Scottish middle-class accents in addition to RP, they found a preference for the ingroup (Scottish) varieties across solidarity and status dimensions.

Although most work in this field has centered around a three-way connection between language, group membership and personal qualities, we should not assume that a conscious knowledge of a specific group membership forms the crux of this picture (Preston and Robinson 2005). Supporting this cautionary point is Ladegaard (1998b), who showed that Danish listeners with minimal exposure to English-language

cultures share stereotypes appropriate to the regional and social varieties of English, although their identification rates for the varieties are relatively low. This suggests that through exposure to performances of English in media, they have incorporated such stereotypes while bypassing knowledge of the actual regions or countries involved. Williams *et al.* (1999) provide further evidence, documenting that while teens in Wales showed very poor performance recognizing the specific dialect regions of other teens, they nonetheless grouped speakers from each region together on a scales of Welshness. This seems to indicate that although the listeners did not have conscious knowledge of the dialect regions, their social calculations took regional differences into account.

These findings regarding the relative perceptions of status and a social attractiveness for different varieties are interesting, but they represent only a small fraction of the meanings and associations that speakers may hold with respect to language. Deciding what kinds of questions to ask participants is one of the great challenges of this type of research.

Expanding the range of reactions

A crucial aspect of the MGT is the selection of the questions to ask. The most common approach is to present listeners with a set of rating tasks, allowing them to indicate (on a scale of one to six, or one to seven, etc.), how educated/uneducated or friendly/shy the speaker sounds (a **semantic differential**) or alternatively how strongly the listener agrees with a statement like “I would want to have this person over for dinner” (a **Likert-type item**). Even within this restricted framework, there is a wide range of variation for selecting the qualities to be evaluated.

Zahn and Hopper (1985) conducted a thorough overview of the measures that had been used to that point and developed a comprehensive version, which they named the Speech Evaluation Instrument (SEI). One of the more remarkable aspects of this literature is the ease with which researchers all over the world have induced subjects to rate disembodied voices on an amazing array of qualities including not only intelligence, correctness of speech or friendliness, but also height, physical attractiveness and even metaphorical qualities such as sweet/sour. Despite this variety, Zahn and

Hopper note that in most of the existing work the variation in responses can be summarized by a few core dimensions.

Virtually every study has found one factor involving intelligence, although depending on the other available attributes it also encompasses socio-economic status, education, competence, correctness of language, height and good looks. Likewise, most studies have also found a factor called likeability, solidarity, trustworthiness or aesthetic quality. Many researchers have also identified factors related to dynamism, confidence or self-assurance. Zahn and Hopper also note a few studies which, looking at standard vs. nonstandard accents, have identified nonstandardness or ethnicity as a major factor, for example Williams *et al.* (1976). The similarity in the names of the factors through the literature should not be taken as an indication of a deeper consistency, however, as the details of which actual qualities fall into each factor differ from study to study.

While Zahn and Hopper had an eye specifically on developing an approach to listener evaluations of linguistic diversity (specifically regional and ethnic diversity), they suggest in their discussion that it may be general enough to be used in other speech evaluation tasks. The generality of their set of attributes is clearly a strength in some ways, for example it allows for easier accumulation of results across studies. At the same time, it is also be a weakness, as it will fail to detect meaning differences that are framed entirely in local meanings. For this reason, some researchers have chosen to conduct pilot studies to discover the appropriate terminology for a specific setting (Lambert *et al.* 1960; Fraser 1973; Williams *et al.* 1976), or adopted other methods for insuring the meaningfulness of their evaluation instrument (Wölck 1973).

In addition to having listeners evaluate speakers along individual qualities like intelligence, some researchers have explored other qualities, such as the degree of control listeners believe the speaker to have over their speaking style (Bradac and Wisegarver 1984). Others have investigated the impact of speakers' linguistic traits on listeners' moods and the degree to which listening to the speaker made particular aspects of their own identities salient (Cargile and Giles 1997). In some cases, researchers have asked participants not only to indicate their evaluations, but also to separately describe how confident they are as to their accuracy (Edwards 1979).

When the message content includes a purportedly real-life narrative, it is possible to also ask listeners questions about the specific events involved, for example how responsible the speaker was for an automobile accident, or how guilty they seem to feel about it (Giles *et al.* 1990). Working with child respondents, Giles *et al.* (1983a) had them draw pictures of the speakers, to investigate whether their status perceptions would be reflected in the relative sizes of the people in the drawings. Persoons (1988) asked Flemish high-school students to describe the hypothetical responses of the majority outgroup (francophone Walloons) to Flemish speakers with particular linguistic choices and histories.

In a handful of cases, researchers have gone beyond the evaluation paradigm to investigate actual interactional strategies in response to experimental stimuli. Work in Accommodation Theory has been particularly successful in developing ways of eliciting actual responses, by manipulating real interactions and noting in which cases participants converge or diverge linguistically (Putman and Jr. 1984; Bourhis 1984). For example, speakers of minority languages such as Flemish or Welsh are more likely to diverge into their own varieties when a speaker of a standard variety makes an explicit challenge to their ethnolinguistic identity (Bourhis *et al.* 1979). Bourhis and Giles (1976) developed another approach for investigating behaviors in response to linguistic attitudes, by using different guises to request moviegoers to fill out a questionnaire regarding in a given theater (also reported in Giles and Bourhis (1976)). This revealed patterns not only overall, but also in the ways that people attending different kinds of events responded differently. This methodology was repeated with intriguing results in Denmark (Kristiansen and Giles 1992).

A further step in the direction of naturally occurring interactions is to investigate the correlations between perceptions and linguistic traits in spontaneous interactions (Scherer 1979). Although this limits the degree of causality which can be inferred from the results, it provides a different perspective which may be combined with more experimental approaches either in a larger research program or within a single study.

Garner and Rubin (1986) investigated linguistic behavior, having a Black researcher interview Black attorneys in the South and deliberately introducing style shifting part way through the interview. As it turned out, this manipulation was

redundant as all but one participant initiated style shifting on their own. Willemyns *et al.* (1997) carried out a very similar study, showing that job applicants in Australia converged towards an interviewer with a broad accent, but not towards one with a cultivated accent. Natale (1975) investigated the degree to which speakers converged towards interlocutors with respect to vocal intensity by artificially manipulating the intensity of the interviewer's utterances during the interaction. Moving away from linguistic responses, Riches and Foddy (1989) used a group pattern-matching task to see how ethnic identity, signaled by accent, affected participants' willingness to accept input from a fictional "partner". Bishop (1979) used actual confederates to examine the influence of race and language variety on participants' perceptions and actual behaviors in a face-to-face simulated interview. Although the study required intensive training of the confederates to attempt to control their behavior across subjects, it revealed some fascinating results, including an effect of race and dialect on the relationship between overt ratings of liking/willing to be friends with the interlocutor and nonverbal behavior such as body position and eye contact. For these White subjects, these measures were positively correlated when interacting with the White confederate, not correlated when interacting with a standard-speaking Black confederate and negative when interacting with a Black confederate speaking Black English.

Researchers in more applied areas have used a variety of evaluation techniques to explore the factors involved in hostile speech, such as sexual harassment (Kinney 2003) and racial insults (Leets and Giles 1997; Leets 2001; Leets 2003). Social psychologists examining the effects of prejudice against various minority groups have developed a range of tools for investigating links between prejudice (as measured by attitude questionnaires and the like) and behavior. One such example is the lost email technique, investigating how frequently individuals with more or less prejudiced overt opinions returned a "misaddressed" e-mail to the senders, based on whether the e-mail contained good or bad news and whether the intended recipient appeared to be a member of a minority group (Bushman and Bonacci 2004).

The evaluation task is not the only area in which we benefit from an increase of complexity. Earlier research using the matched guise technique aimed to divorce

the task as much as possible from the extralinguistic context, including the speaking situation and the content of the message. Researchers have increasingly shifted their perspective to one which acknowledges the centrality of context. This theoretical shift has likewise been accompanied by some fascinating methodological innovations.

Increasing the complexity of the judging situation

One of the important lessons to emerge from the language attitudes literature is the necessity of dealing with the context of speaking (Lee 1971). This has not always been a concern in the same way, as the early studies strove for “neutral” message content and gave no contextual information to their subjects. Researchers rapidly came to realize that conversational context is always a variable and if it is not provided, subjects will fill it in, explicitly or implicitly, in the process of evaluation (see Bradac *et al.* (2001) for a discussion of this). What exactly constitutes the context of a given utterance is complex, since the information influencing listeners’ evaluations shifts as new social moves are made and as a given interaction develops, “different language behaviors are afforded different salience” (Cargile *et al.* 1994:216).

The current state of the field deals with context in much broader terms however, for example contrasting conversations set at school with those from a social interaction (Callan *et al.* 1983; Creber and Giles 1983; Gallois *et al.* 1984), in a social interaction with an employment interview (Street and Brady 1982; Jr. *et al.* 1983) or other formal/informal situations (Hogg *et al.* 1984; White *et al.* 1998). Much of the research into the effect of context has looked at the complex ways that context interacts with other factors. Cargile (1997) found an important contextual influence on the impact of accented English, showing that listeners graded speakers in a job interview setting equally well regardless of whether they had an American or a Chinese accent, while if the same recordings were presented as excerpted from a lecture, the Chinese accented guise was significantly downgraded.

Context may be manipulated to eliminate the effects of linguistic variables as well, for example, speech rate generally influences ratings of competence in a linear way with slower speakers rated as less competent. But this effect disappears if the speakers

are introduced as explaining a difficult topic to an audience of beginners (Brown *et al.* 1985). Instead of manipulating the context, listeners may in other cases be asked to describe which contexts are appropriate for specific ways of speaking (Taylor and Clément 1974; Hoover 1978).

Much of the work on context has approached it in the broad sense as the setting in which the speech is taking place. Some suggestive research has also been done on more specific aspects of context, such as the content of the message being evaluated (Ray *et al.* 1991). Cargile and Giles (1998) examined the influence of message content, looking at the interaction between accent (Japanese- vs. American-accented English) and the “aggression” of the message, specifically whether the speaker explicitly criticized US policies. They found that there was an interaction between accent and message content on the ratings of dynamism. Smith and Bailey (1980) demonstrated that both message content and speech activity (reading or speaking spontaneously) influenced speaker perceptions significantly.

Other aspects of context include preceding message(s) (Bradac *et al.* 1977; White and Li 1991), or the information provided to listeners regarding the situation, for example whether listeners are informed that the speaker is of high or low social and educational status (Bradac and Wisegarver 1984), or that the elderly addressee resides in a nursing home or in the community (La Tourette and Meeks 2000). The friendliness or likeability of the speaker also has an impact on the evaluation of other aspects, for example their competence, even though listeners may not be aware of the effect (Nisbett and Wilson 1977). Other mood-influencing aspects of the interaction also affect ratings of speakers, including artificially introduced white noise which interferes with a communication task (Sebastian *et al.* 1980; Ryan and Laurie 1990).

Another way of thinking about context is the larger context of the speakers, listeners and the study itself. For example, researchers into issues of fairness in education have investigated the role of linguistic variation in teacher perceptions of students (Seligman *et al.* 1972; Williams 1973; Edwards 1979; Slaughter 1982; Yamamoto and Hargrove 1982) or student perceptions of teachers (Edwards 1977; Gill 1994). Work in education is particularly relevant given the very real social impacts of language bias in the educational context and the potential for research and

awareness of these issues to mitigate those effects (see Bradac and Giles (1991) for a discussion). Another context with very real applied consequences is that of patronizing or accommodative speech used by caregivers towards elders (Edwards and Noller 1993; La Tourette and Meeks 2000). Others have investigated the role of linguistic traits on perceptions in court, both evaluating the impact of regional or non-standard accents on how guilty a suspect sounds (Seggie 1983; Dixon *et al.* 1994; Dixon *et al.* 2002) and the influence of powerful/powerless speech styles on the credibility of witnesses (Lind and O'Barr 1979; Hosman 1989; Nigro *et al.* 1989; Schmidt and Brigham 1996; Hosman 1997; Ruva and Bryant 1998; Smith *et al.* 1998; Sparks *et al.* 1998). A handful of studies have addressed evaluation of medical messages, i.e. doctors explaining illness or treatment regimens to patients (Ray *et al.* 1991). Because of the natural opportunities for evaluation, job interviews have also been popular contexts to explore (Hopper and Williams 1973; Shuy 1973; Zerda and Hopper 1979; Kalin *et al.* 1980; Ball *et al.* 1984). A large body of work in social psychology is devoted to understanding the factors which influence the success or failure of persuasive messages, including some linguistic or paralinguistic cues (Miller and Hewgill 1964; Sereno and Hawkins 1967; Pearce 1971; Sparks *et al.* 1998).

Although they are not usually considered as such, the listeners themselves also form a part of the context of evaluation. For example, whether listeners are participants in an interaction or merely observers has been shown to influence their responses. Street (1985) had student interviewers and interviewees rate each other and compared these ratings to those made of both by observers listening to the tapes afterwards. He found no correlation between the ratings of participants and those of observers and in some cases found that they displayed opposite preferences. For example, interviewers rated those interviewees with longer response latencies and slower speech rates as more competent, while observers had the opposite response. Although not dealing directly with language, Harkness *et al.* (1985) supports this point by documenting that participants in their study judged a fictional person differently and used more complex reasoning to form their judgments when they believed that they would be interacting with that person on a regular basis.

Researchers have also investigated the role of individual listener traits, not only

their ethnic or other demographic memberships, such as age (Ball 1983; Paltridge and Giles 1984), sex (Paltridge and Giles 1984) and regional background (Paltridge and Giles 1984), but also more specific qualities. For example, Aune and Kikuchi (1993) demonstrated that both the actual and perceived similarity between speaker and listener in terms of language intensity correlated with, among other things, listeners' evaluations of the speakers' credibility and how likely the listener was to agree with the speaker. Edwards and Noller (1993) found that elderly women rated potentially patronizing communication strategies as less patronizing and more respectful than nursing and psychology students did. It is also possible to investigate the connections between specific evaluation behaviors and more abstract listener qualities, as measured by various profiling questionnaires, such as racial bias (Robinson 1996) or "African self-consciousness" (White *et al.* 1998).

Because these studies are usually carried out by playing listeners pre-recorded interactions, the speaker's interlocutor also forms an important part of the context, distinct from the listener. Researchers working in the framework of Accomodation Theory (discussed in Chapter 1) have used the MGT to explore listener perceptions towards shifts in languages or varieties relative to an interlocutor. For example, Australia listeners rated a job applicant as particularly flippant when they heard the applicant maintain or enhance a "broad" accent when interacting with a job interviewer who had a "refined" accent (Ball *et al.* 1984). Similar work has also been done looking at interactions between French and English speaking Canadians (Giles *et al.* 1973; Genesee and Bourhis 1988), English speakers and Cape Afrikaans speakers in South Africa (Dixon *et al.* 1994) and different ethnolinguistic groups in Switzerland (Doise *et al.* 1976). Convergence or divergence can also occur with respect to smaller variables such as lexical diversity (Bradac and Giles 1988), or speech rate, utterance duration and response latency (Street 1982).

These contextual factors represent some of the most exciting developments of the field. As we deepen our understanding of the ways that different factors in the linguistic interaction influence each other, we deepen our understanding of the social structures built by language. However, much of this work takes the linguistic material itself to some extent for granted, identifying linguistic changes in broad and sometimes

imprecise terms. This is where sociolinguists come in, with our passion for linguistic detail and for discovering exactly what aspects of a linguistic performance are relevant for particular social constructs.

Sociolinguistic approaches

Although sociolinguists have participated in the language attitudes literature, they have commonly adopted very different perspectives and approaches than many of those described above. Sociolinguists and social psychologists often have different conceptions of what constitutes an interesting question about language. In some cases, sociolinguists have used perception experiments similar to the MGT to investigate how listeners are able to identify speakers' regional origin or other information. These studies have investigated not only whether listeners are able to identify regional variation, but also frequently examine specific linguistic correlates of given identifications.

In other cases, sociolinguists have conducted more typical evaluation studies, but sought to identify the specific linguistic traits which trigger the evaluations. One example is Labov (1966), who included in his comprehensive treatment of variation in New York an evaluation study of the main vocalic variables in his study. Instead of using the MGT, he took recordings of a number of speakers reading the same passage aloud. The passage itself was constructed so as to showcase each variable in a different paragraph. The listeners heard a different speaker read each sentence in a given paragraph and were asked to rate the speakers in terms of the highest profession they could hold, given their speech. Because of the manipulation of the passage, each sentence elicited evaluations of each speaker's value on a single variable.

Similarly Rickford (1985) used the matched guise technique to elicit reactions in Guyana to samples of speech representing three points along the Creole continuum, such that speakers along different points used the appropriate percentages of the relevant variables. Rickford found that while speakers tended to agree on the relative socioeconomic status of speakers using basilectal, mesolectal and acrolectal forms, their solidarity ratings tended to favor the speakers most like themselves (Rickford

1985:152) (see also Rickford and Traugott (1985), which further incorporates media analysis).

Another area of research is the ways social information influences the processing of sounds, discussed in Chapter 1. This work does not follow the methodological paradigm of the MGT, as it focuses on the influence of extralinguistic information on linguistic perception, rather than the influence of the variant on the social evaluations. Research in this paradigm has explored the influence of visual cues (McGurk and Macdonald 1976), gender (Johnson *et al.* 1999; Strand 1999) and nationality (Niedzielski 2001; Niedzielski 1999). It has also looked at the perceptual aspects of bilingualism (Elman *et al.* 1977).

Other linguists have employed a typical MGT approach to investigate evaluative reactions to specific sociolinguistic variables. With the advent of software allowing for more natural-sounding manipulated recordings, it has become possible to create minimally paired stimuli to elicit listener regional identifications, for example. Plichta and Preston (2005) showed that listeners are capable of aligning a linguistic cline with a geographic one, even when they believe that they cannot. Similarly, Fridland *et al.* (2004) used resynthesis to investigate connections between perceived Southern accents and a range of vowels. Other perceptual work on variation has examined quotatives such as *like* and *all* (Buchstaller 2003; Buchstaller In Press; Buchstaller To Appear).

The literature on gay speech, discussed in detail in Chapter 5, has relied almost exclusively on listener perceptions. They have looked for correlates of listeners identifying male voices as gay, examining pitch (Terango 1966; Travis 1981), pitch variation (Terango 1966; Lerman and Damste 1969; Travis 1981; Levon 2005a; Levon 2005b), sibilant duration (Crist 1997; Linville 1998; Levon 2005b), sibilant peak frequency (Linville 1998) and overall size of vowel space (Pierrehumbert *et al.* 2004), among others.

As the linguistic units of analysis become more detailed, we begin to stretch the notion of “attitude” to its limit. It seems perfectly natural to talk of a listener’s attitude towards French, or towards a Southern accent. It is somewhat less so to speak of their attitude towards the length of /s/ in an utterance. If we are to successfully marry the fields of language attitudes and variation, a better understanding of the

construct attitude is needed.

Understanding attitudes

As Bradac *et al.* (2001) discuss, attitudes towards a group of people or a particular linguistic performance form part of a larger recursive system of interaction through which speakers conduct their daily business. In the course of this process, attitudes are constantly formed, shared, acted upon, reacted to and reshaped. The attitudes studied in a given situation are affected by a range of things, including the listeners' goals, mood and available knowledge (Cargile *et al.* 1994). The study of attitudes in psychology extends far beyond the study of attitudes about language and psychologists are still in the process of reconciling the established methodologies with appropriately sophisticated theoretical models (Haddock and Maio 2004a).

As sociolinguists, we are faced with the challenge of not only comprehending the psychosocial significance of a listener's overall "attitude" towards a speaker, but also with merging this understanding into our existing models of the social meaning of linguistic choices. This study documents a connection between the use of a given variant of (ING) in a specific situation and a change in the rating of a listener on a list of labelled scales. In social psychology terms, this is establishing information about the relative attitudes of the listener(s) to the two linguistic styles presented. This opens the question of how to translate that information into knowledge about the specific variables themselves and what the theoretical relationship is between the behavior changes registered by the measure and the concepts of "social meaning" or "indexing" discussed in Chapter 1. This bridge is not yet built, but its construction represents the central theoretical work of the research program of which this dissertation is a part.

3.2 Methodological choices

The current study draws on a range of techniques within the literature for its methods. Because it pursues a new kind of information, it combines these existing methods in

a new way. The most important methodological elements of the study are: the use of digitally manipulated stimuli, the use of speech from interviews rather than reading passages, the inclusion of multiple samples from each speaker and the collection of both open ended interview data and categorical survey data. Before going over the methods of my study in detail, I will touch on each of these tools to mention its previous uses and discuss its importance.

Any sociolinguist with an inclination toward experimental methods will dream of being able to perfectly manipulate a speech sample for any variable and create believable results. While we are not yet at that level of technological bliss, we have come so far in even the past 10-15 years that it may seem like we have reached it. Thirty years ago, it was possible to use technology to change the rate, pitch and pitch variation of a sample of speech (Brown *et al.* 1973; Brown *et al.* 1974; Smith *et al.* 1975; Apple *et al.* 1979). Today, we can use available software to alter the formants of a vowel (Graff *et al.* 1986; Fridland *et al.* 2004; Plichta and Preston 2005). While the current technology cannot yet transform an /n/ to a /ŋ/, it can allow for a “cut-and-paste” approach and still produce a natural-sounding token (described in detail below).

The benefit of such manipulation is in the precise control it gives over the behavior to be evaluated. Barring synthetic manipulation, the usual method of constructing matched stimuli is to have speakers (naturally bilingual or bidialectal or occasionally professional actors) perform each version, consciously shifting the variables as they speak. This may be straightforward when investigating attitudes towards whole languages or varieties. When exploring more “molecular” variables (Scherer 1979), it becomes difficult using this method to ensure that only the variable in question is changed. We understand very little about how different variables interact in the perception process, which means that when judging the success of such manipulations, it is unclear how far to trust our perceptions of whether other aspects of the performance have changed. We can still gather interesting and relevant information regarding the larger perceptual consequences of, say, fast or slow speech. What we cannot do is be sure what the precise linguistic triggers of this perceptual change are.

Another advantage of digital manipulation is that it opens the doors to using a host

of different speakers and speaking situations, by eliminating the need for conscious control on the part of the speaker. It is possible to use naive speakers or speakers of only one variety and to create speech samples from spontaneous or even completely naturally occurring speech. Although much of the literature on the matched guise technique has used read speech, there have been exceptions (d'Anglejan and Tucker 1973; Palmer 1973; Wölck 1973; Apple *et al.* 1979; Huygens and Vaughan 1983; Mulac *et al.* 1985; Graff *et al.* 1986). There have also been the other attempts to solve the problem, for example Johnson and Buttny (1982), which took interview transcripts, altered them appropriately, then had the same speakers learn them well enough to read them comfortably.

There are many reasons to prefer spontaneous speech for these purposes. There is a significant body of work documenting the differences between read and spontaneous speech with respect to prosody (Fowler 1988; Bruce and Touati 1992; Dascalu-Jinga 1992; Blaauw 1994; Howell and Kadi-Hanifi 1991; Laan 1997; Guaitella 1999; Hirose and Kawanami 2002). Relative to spontaneous speech, read speech has a different rate (Kowal *et al.* 1975), a different number, length and arrangement of pauses (Kowal *et al.* 1975; Barik 1977; Howell and Kadi-Hanifi 1991; Guaitella 1999), different patterns of shortening for repeated words (Fowler 1988) and different tone boundaries (Howell and Kadi-Hanifi 1991).

Not only do such linguistic differences change the linguistic performance itself, they signal the origin of the speech so that listeners know that it was produced by reading aloud. Listeners show agreement in how to punctuate speech that has been read aloud from a text, but not how to punctuate spontaneously produced speech (Guaitella 1999) and the process of detecting phonemes is influenced by different factors in the two modalities (Mehta and Cutler 1988). Listening to recordings of teachers reading or telling stories to children, listeners could distinguish the two even when the recordings were manipulated to eliminate segmental information (Levin *et al.* 1994). More significantly for the current study, Giles *et al.* (1980) found that the relative intelligence ratings between feminist and non-feminist women's voices were completely reversed between the spontaneous and read speech conditions. Smith and Bailey (1980) demonstrated that listeners' evaluations differed significantly based on

whether the speakers were reading aloud or speaking spontaneously and further found that the specific effects of speech activity differed for different speakers.

When using read speech, most of the research has used linguistic variation primarily as a foil for getting at covert judgments of social groups. As a result, both the researchers and the listeners are assumed to subscribe to the ideology that a person's essential characteristics come through in their speech regardless of the context of speaking. In this research, I have shifted the focus away from beliefs about personality and looked more widely at the interactions of linguistic variables, message content and idiosyncratic speaker traits. To do this, it was necessary to use spontaneous speech to create the recordings.

One of the potential drawbacks to using spontaneous speech is that it is difficult to control content of the recordings. Unsurprisingly, the content of what speakers say impacts the judgments others make about them. This has been an ongoing issue in the literature but for the most part work has been devoted to minimizing its influence, primarily by creating "neutral" content. This construct has been critiqued a number of times (e.g. Agheyisi and Fishman (1970)) and rightly so. Clearly, content and factors like word choice and sentence structure influence the interpretation of the speaker and may influence the role of other variables, including the ones of interest in a given study. To address this issue, I included four recordings from each speaker, in order to begin to explore the role of content as well as provide some variation within the data from each speaker.

The key elements of this method are the stimuli and the evaluation task that the listeners perform. As discussed above, there is a standard collection of qualities which tend to collapse into familiar factors such as status, social attractiveness and trustworthiness. These selections are primarily designed for learning about the fundamental personal differences which divide members of different linguistic groups perceptually. Having a standard set of qualities is useful for building on others' work (Ryan 1979) but it presupposes that the standards are the relevant ones for a given population and stimuli. In this study, I employed open ended group interviews which provide a rich data source in their own right as well as

a set of pilot data for the second stage of surveys. This kind of piloting is unusual but does have a history in the literature (Lambert *et al.* 1960; Fraser 1973; Williams *et al.* 1976), while other studies have developed culturally appropriate and nuanced stimuli without the use of formal piloting (Wölck 1973). Instead of piloting it is also possible to ask listeners to describe their reasoning after completing a more restricted questionnaire, allowing a deeper insight into the evaluation process (Giles *et al.* 1990).

The rest of this chapter will be devoted to describing the methods of the study in detail, beginning with the creation and manipulation of the recordings, then describing the the methods for data collection and finally the statistical analyses.

3.3 Creating the stimuli

As discussed in Chapter 2, the existing literature and the pilot study suggested that there may be regional differences in the use and interpretation of (ING). Accordingly, I sought participants from two distinct areas in the U.S., one from within the South (North Carolina, specifically) and one from outside the South (California). Because so much of this experiment was breaking new ground, I used university students as both speakers and listeners since they are easily available. I selected eight speakers in all. Table 3.1 gives the names (pseudonyms) of the speakers, divided by region and sex. Four were students at North Carolina State University, two men and two women,

	Women	Men
North Carolina	Bonnie	Robert
	Tricia	Ivan
West Coast	Elizabeth	Sam
	Valerie	Jason

Table 3.1: Speakers, by region and sex.

all of whom had grown up in North Carolina, in the middle or eastern portions of the state. I chose NCSU due to the large population of in-state students, which increased the chances of finding speakers who were from the area and who had perceptible

Southern accents. As it turned out, only three had perceptible accents, a fact which created interesting complexities in the data. The four West Coast speakers were all students at Stanford, three of whom had grown up in California and one of whom was from Seattle. All eight speakers were White, as the literature suggests that race has an effect on (ING) use and I did not have the resources necessary to incorporate race as a variable.

Gathering material

To make the recordings, I met with each speaker twice. At the first meeting, I explained the overall structure of the study and told them I would be manipulating excerpts of their speech and playing them for others to evaluate. I did not tell them what linguistic features I would be changing. We then conducted a sociolinguistic interview, approximately one hour in duration. The first part of the interview focused on schoolwork or other work experiences and the second part on their recreational activities and/or family life. The questions were designed to find topics of interest to the speakers and expand on those to obtain fluent, comfortable speech. I attempted to put speakers at ease and avoid awkward speech but did not try to elicit a particular style of speech or obtain the speakers' vernacular.

After each interview, I transcribed the resulting tape, then met again with the speaker. I explained in more detail the point of the study, including the variable to be manipulated. We went through the transcript and I asked them to produce alternate tokens for each instance of (ING) from the original interview. For instance, I would indicate a point in the transcript, for example "I'm planning on going to grad school." I would play that excerpt from the original interview and ask the speaker to capture the speed and intonation of the original as much as possible. They would then produce two tokens "I'm planning on going to grad school." and "I'm plannin' on goin' to grad school." In this way, I collected alternate tokens for every occurrence of (ING) that was likely to be of use in creating the final recordings.

Selecting excerpts

Once the raw material was available, the next step was to select short clips of each interview to use in the study. In order to explore the effects of message content and topic, I chose four passages from each speaker: two about school or work and two on recreational or family topics. Each excerpt was a single uninterrupted turn, with only the speaker's voice audible on the tape.

After experimenting with multiple possibilities for the length of the recording, I determined that recordings of approximately 15 seconds in length would provide enough material to give listeners adequate information and enough space for multiple tokens of (ING) to appear. On the other hand, they would be short enough to enable listeners to hear and judge all eight speakers without undue fatigue.

Although the aim was to find recordings 15 seconds in length, I made it a priority to select recordings that were relatively coherent, with natural breaks in content and intonation. The recordings ranged in length from 10 seconds to just over 20 seconds. In a couple of cases, it was impossible to find an excerpt of appropriate length with sufficient tokens of (ING). In these cases, I did a small amount of manipulating. For example, for one excerpt I removed a portion of the recording containing my own minimal response, so as to create a longer uninterrupted turn for the speaker. In another, I joined two turns together into a single utterance, eliminating an intervening question.

The number of tokens of (ING) varies in each recording, ranging from two to six tokens. The majority of the recordings had three or four. For the transcripts of each recording, see Appendix A.

Manipulating the variables

I used the software package Praat to digitally manipulate the selections, creating minimal pairs of recordings. I spliced in both *-in* and *-ing* variants, to minimize confounding differences introduced by the manipulation. Although manipulating both sets did not ensure identical recordings, it made it more likely that differences based on artifacts of the manipulation process would be minor and limited in scope, rather

than consistent across all tokens.

In order to minimize variation across token pairs, I created a routine for the manipulation, departing from it only when necessary to create believable tokens. For most of the tokens, I had matched alternatives available, tokens of *-ing* and *-in* in matching phonetic environments created in the follow-up sessions with each speaker. In a handful of cases, the alternates did not exist or were not usable. When this happened, I took alternates for other tokens that were phonologically similar, ideally tokens occurring in the same lexical item elsewhere in the interview.

The first step to creating the minimal pairs was to open a copy of the original excerpt from the interview and adjust the window to include the token to be manipulated. I then opened files containing the alternates, one of the speaker saying *-in* and one of the speaker saying *-ing* in similar phonetic environments as the original. After selecting and opening the three excerpts, the next step was to select exactly the phonetic material to be cut and pasted. Figure 3.1 shows an image of the window used to do this, giving both the spectrogram and the wave form. It was important to

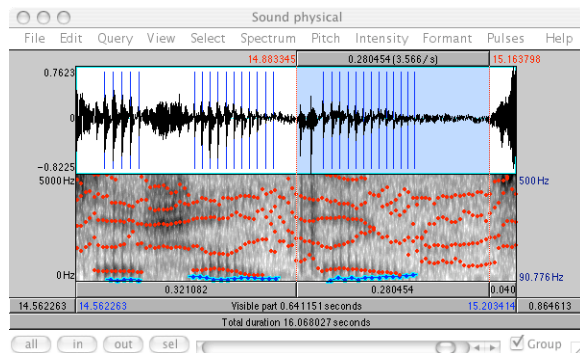


Figure 3.1: Selecting the appropriate material: *thing out of something*.

ensure that anything which helped to identify the nasal was included. This was useful in making certain that the most important parts of the alternate tokens were substituted in, but much more important for removing any signals which would carry cues to the original token. Apart from that constraint, I selected the minimum amount of material possible, ideally just the vowel and nasal of the token itself. In selecting the point for splicing, I looked for either a pause, a stop closure or a point within a stable sonorant. If it was necessary to make the change point in the middle of a sonorant, I

used Praat’s formant tracking feature to find points in each alternate and the original which were similarly located in time and matching roughly in their formant values.

Once the boundaries of the three tokens were selected, I needed to adjust the alternates with respect to pitch, intensity and length in order to match the original as much as possible. As discussed above, the characteristics of spontaneous speech are quite different from that which is read aloud or otherwise recited, in addition to the fact that individual tokens may vary a great deal in any case. As a result, it was usually necessary to adjust the alternate tokens so that their intensity, length and pitch matched the original utterance. To accomplish this, I used Praat’s facilities for examining and altering these qualities. Adjusting the intensity was the first step, multiplying the excerpt so as to match the amplitude to the original. After adjusting the intensity to the appropriate level, I altered the pitch. Praat allows a user to create a “pitch tier” from a sound file, a charting of calculated pitch over time, shown in Figure 3.2. It is possible to then apply this sequence to a different sound file, adjusting the pitch of the new file to match the pitch track of the original. I excerpted the pitch track from the original and used it to resynthesize the alternate tokens.

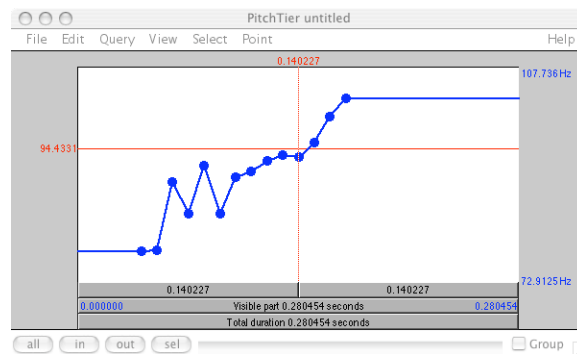


Figure 3.2: The pitch track.

I also altered the length of the alternates, usually to match the length of the original. In most cases, I used a straightforward multiplication provided by Praat, where I gave the ratio of the length change that I wanted and the tokens was evenly recalculated for the new length. In some cases, the relative lengths of the vowel and the nasal were too disparate between the tokens. When this happened, I use Praat’s manipulation feature, as shown in Figure 3.3. This has a length track which

allows the user to create change points and adjust individual portions of an excerpt for length. Using this feature, I was able to alter the lengths of the vowel and nasal independently.

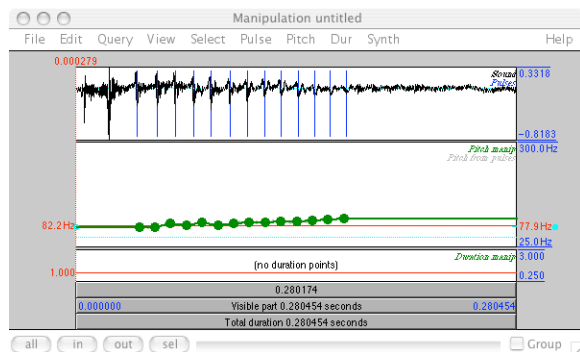


Figure 3.3: The manipulation window.

In some cases, either the length or of the pitch of the two alternate tokens diverged so remarkably from the original that altering them both to match it resulted in a bizarre or impossible utterance. When this happened I placed priority on matching the manipulated alternates to each other and on obtaining a natural sound. In several cases, this meant that the newly manipulated recordings featured (ING) in a more prominent and easily heard setting than the original token from the interview. In a very real way, I was manipulating the social content of the recording. I was careful to maintain concurrence between the two manipulated versions, so that the new alternates would match.

The issue of how similar the two members of the minimal pair needed to be was a tricky one. (ING) has been well studied as a sociolinguistic variable, but little work has been done concerning the specific phonetic attributes of the variants and how they differ, if at all. For example, consider the length of the variant. It is not clear whether under normal circumstances there would be regular differences between the two variants with respect to length. If there are such differences is not immediately clear what a minimal pair would be. On one hand, using the same lengths and pitches would create the most strictly similar tokens. On the other, if the normative length for the two variants differ, using identical lengths shifts the comparison to one between a typical token of one variant and an atypical token of another.

A thorough examination of these issues was beyond the scope of the study. From working with the variables, I suspect that there are no regular pitch differences, but that length may be an important issue. Tokens of *-in* in the interviews and the alternates tended to be shorter than their corresponding *-ing* examples. Problems with obtaining a natural sounding token occurred when examples of *-in* were lengthened too much. I attempted to match the lengths but paid special attention to the influence of length on the naturalness of the recordings, avoiding tokens which sounded emphatic or marked in some way. There also seemed to be some regularities with respect to the quality of the vowels found in the two variants. However, manipulation of this quality was beyond my capability, as a result I left the effect of this for later study.

Through the course of the work, I relied on my own ear to judge naturalness and identifiability. Once all of the recordings had been created, I ran a short pilot to confirm that the tokens were identifiable as the variants intended. I took excerpts from the manipulated recordings, including a couple of words on each side of each token. Listeners were asked to identify whether each token was *-in* or *-ing* and to indicate whether the recording sounded strange. Each token was heard by at least five listeners and tokens with more than two misclassifications or reports of strange sound were remanipulated and repiloted.

3.4 Collecting listener reactions

Earlier in the chapter, I discussed the importance of getting both open-ended reactions in addition to more restricted responses. I used group interviews and a web based survey respectively, to accomplish these. Because I conducted the interviews first, I was able to use them both as data in their own right and as a pilot for the survey. Having information about what terms speakers use spontaneously to describe the speakers helped me to select the terms used in the survey and to design its structure. Likewise, information about the ideologies and beliefs interview participants had regarding (ING) suggested areas of meaning to investigate more rigorously in the survey.

Gathering intuitions: group interviews

Once the manipulated recordings were finished and tested, the next step was to collect open ended responses. In conducting the group interviews I had two goals. The first was to determine what the general reactions were to the speakers in each recording and what words were used spontaneously to describe them. I also wanted to gather native speaker intuitions regarding (ING) and its effect on these particular utterances. In the first part of the interview, listeners heard individual recordings from four of the eight speakers. In the second part, I played the same recordings again but in their minimal pairs, asking the listeners to comment explicitly on how (ING) changed their perceptions.

In order to widen the range of geographic regions among the participants, I conducted interviews at three schools: Stanford, Duke and the University of North Carolina at Chapel Hill. The latter two schools were selected for two reasons: their location in North Carolina, the same area where the Southern speakers were from and their relatively high status. In early interviews at Stanford and North Carolina State University, I determined that the evaluations were likely to be influenced by the educational background of the speakers, even (or perhaps particularly) by the prestige of the school they were attending. At all three schools, subjects were solicited through fliers and advertisements in campus newspapers. In the case of the North Carolina schools, the interviews were carried out during the summer, among students who had stayed for summer classes, or who were working on campus.

In order to keep the task to a reasonable length, I divided the recordings by gender and content. One group heard only recordings of men discussing work related topics, another heard men discussing recreational topics and so on. I began by explaining the purpose of the study and going over the consent form. I did not identify the particular variable we would be discussing but explained that I was trying to learn about how small differences in the way we speak influence how we are perceived. I explained that we would be listening to a series of recordings and that I wanted their general impressions as to who the people were. I stressed that I was interested in finding out what they were able to tell about the people from listening to the recordings and what they were not able to say, so they should not invent answers if they had no opinion.

I also emphasized the importance of hearing from everyone. The groups ranged in number from one-on-one interviews up to groups with six participants. The bulk of the interviews were with two or three participants.

For the first portion of the interview, I went through each of the four speakers (all men or all women) one at a time, playing one recording from each. After each recording, I asked the participants to tell me what came to mind in response to the recordings and what they were able to tell me about the speaker. After a round of answers to this question, I asked whether the speaker seemed like he/she was competent, or good at what they did. Next, I asked whether the speaker sounded like someone the participants could see themselves being friends with. After discussing the speakers themselves, I asked what the speaking situation was and who the addressee might be. If no one had already mentioned where the speaker was from, I asked about this last. After having gone through the four speakers once, we listened to a second recording from each speaker in which they said different things on the same kind of topic, either work or recreation. The second recording for each speaker contain a different variant of (ING) than the first, but this was not brought to the attention of the participants. I asked participants to discuss if and how the second recording changed their impression of the speaker and whether the context seemed to be the same as the first recording.

Transitioning to the second half of the interview, I explained the goal of my study in more detail, introducing the variable (ING). I explained that we would be listening to recordings in pairs and that I wanted to know how the change in (ING) influenced their perception of the speaker, if at all. When playing the recording pairs, I identified which speaker we were about to listen to both by name and with a short description identifying which recording we would hear by content. For example, “And now here’s Tricia again. She was the one that was talking about growing cucumbers before. In the recording we heard already, she said *-ing*.” I played the recording that they had heard originally and explicitly identified which variant they had heard. I introduced the second member of the pair by explaining that it would be the same as the first, only with the other variant, again saying whether it was *-in* or *-ing*. In this way, I went through the pairs for all eight recordings from the first half of the interview. The

whole interview took approximately one hour. Before ending the interview, I asked participants whether they had any further comments about any of the speakers, then asked them if they had questions for me. In all, I conducted and analyzed data from 20 interviews consisting of 55 participants.

Testing intuitions: matched guise survey

The two data collection phases collected very different kinds of information. The interviews elicited explicit ideologies involving (ING) and reactions to the recordings overall. The first portion of the interview provided important information on which aspects of the recordings listeners were likely to focus on and what kinds of social assumptions and deductions they made. The second portion of the interview gave insight into the ideologies that listeners held about the role of variation in general and the meaning of (ING) in particular. The goal of the survey was to investigate covert reactions to (ING), testing which meanings are actually influenced in online interpretation of speakers. Listeners were not directed towards (ING) or any other specific linguistic attribute and each listener heard only one recording from each speaker.

The first step in developing the experimental design for the survey was analyzing the interview data for the descriptions to use in the survey. In selecting terms used in the interviews, I used two primary criteria. I sought the most central characterizations of the speakers, on the basis of which were most frequently mentioned, mentioned early on in the interviews and frequently presupposed. I also selected terms which were frequently referenced in the discussions of (ING), either explicitly or implicitly. Although I drew primarily on the interview data, I also combined these with concepts drawn from the production literature on (ING), discussed in Chapter 2 and terms used in the language attitudes literature.

Listeners in the survey heard one recording each from all eight speakers. The speakers were arranged in random order and each listener was randomly assigned to two possible orders, one the exact reverse of the other. I chose to have only two alternate orders in order to be able to investigate the effect of order, something which

would not be possible if each listener heard a different sequence. Each listener was randomly assigned to one of two (ING) guise conditions, in which they heard one member of each gender/region pair (e.g. Southern women) say *-ing* and the other say *-in*. In condition one, listeners heard Bonnie, one of the Southern women, saying *-ing*, and Tricia, the other Southern woman, saying *-in*. In condition two, the assignments were reversed. There were also four groups based on content, two with work topics and two that were about recreation. These did not change, so for example, Bonnie's recording about playing volleyball was always in the same group as Tricia's recording about hiking. This introduces a confounding variable since the preceding sentences undoubtedly influence the interpretation of a recording. The interactions between individual recordings and those preceding would be interesting but are beyond the scope of the current study. Examining the effect of what order the recordings were played in does begin to address this issue.

At the beginning of the survey, participants filled out a brief demographic questionnaire. It asked for their age, the school they attended, the countries or states they had lived in and their racial or ethnic identity¹.

After this preliminary page, listeners responded to eight survey pages, one for each recording. The survey form had six sections. The first was a set of adjective scales ranging from one to six, shown in Figure 3.4. Note that the last item was labeled either masculine or feminine according to the gender of the speaker. Although these qualities are by no means mirror images of each other, time constraints necessitated that I limit myself to only one per speaker.

The image shows a screenshot of a web browser window titled "Survey". The content of the window is a list of seven adjective scales, each with six radio buttons for selection. The scales are:

- Casual Formal
- Intelligent Not Intelligent
- Educated Not Educated
- Very Accented Not At All Accented
- Very Shy Very Outgoing
- Talking Very Slow Talking Very Fast
- Not At All Masculine Very Masculine

Figure 3.4: Adjective scales.

¹Appendix B shows examples of the pages used to collect responses to the speakers.

In the second section, listeners were asked to indicate how old the speaker sounded, given the choice of five checkboxes labeled with different age ranges: *teenager*, *college-age*, *under 30*, *in his/her 30's* and *over 40*. This selection of the age categories may seem a bit odd, until one remembers that it was developed after gathering intuitions from a set of college students. Listeners were required to select at least one age description but could select as many as they chose in any combination.

The next section contained an adjective checklist, listing a range of personal qualities, such as *lazy* and *religious* and social categories or professions such as *redneck* and *artist*. After these personal descriptions came questions relating to the context. There was another six point scale, asking how well the speaker knew the addressee, ranging from *best friend* to *stranger*. Another set of checkboxes asked about speech activities such as *joking* and *chatting* and stances such as *bored* and *polite*.

Lastly, the listeners were asked to guess the background of the speaker. Eight options for region were provided: *the South*, *New England*, *the Midwest*, *the West Coast*, *the East Coast*, *the Southwest*, *the North* and *anywhere*. There were also options referring to community type: whether the speaker was from *the city*, *the country* or *the suburbs*. The last set of boxes asked about class background: whether the speaker was from *a working class background*, *a middle class background* or *a wealthy background*. In each of the checkbox sections except for age, one of the boxes labeled *other* provided listeners with the opportunity to add their own descriptors. At the end of the page an open-ended question invited them to add any remaining thoughts they had about the speaker.

The survey itself was administered over the World Wide Web, allowing for a wider geographic distribution on a smaller budget. The length of the survey was adjusted to allow listeners to complete the entire process in approximately 15 minutes. This included reading over the consent form, filling out the demographic questionnaire and completing eight individual web forms, one for each speaker. Lastly, listeners were presented with a page giving options involving incentives and giving them an opportunity to send comments or feedback. Participants were recruited through word of mouth e-mail and classified advertisements in the school papers at Duke and at Stanford. They were offered incentives of \$5, sent either through the postal service,

through the electronic payment system of PayPal, or donated to the Make A Wish Foundation. A total of 124 participants completed the study. An additional 36 began it but failed to finish and their data were removed from the analyses.

3.5 Analysis

The data collected by these procedures created a data set of incredible richness. The complexity of the connections being studied was such that, with the number of participants I had, I was able to only scratch the surface of the interactions and connections present.

Different statistical techniques were needed for examining the three different categories of variables involved: independent variables, checkbox variables and rating variables. The independent variables gave information about which recording was being evaluated (the speaker, the recording itself and) the (ING) variant and about the listener (school, gender, regional background and race). The checkbox variables were binary (or nominal) variables indicating which yes/no attributes the listener had selected (e.g. *articulate*, *artist*). The ratings variables were numerical (or interval level) variables, ranging in number from one to six (e.g. *not all educated/very educated*), indicating the listeners' selections on these scales.

I used the Chi Square test to investigate the influence of the independent variables on the checkbox variables, as well as levels of co-occurrence between the checkbox variables. This test reported the degree to which the proportions of listeners selecting, for example, *articulate* were unexpectedly different among listeners hearing *-in* and listeners hearing *-ing*. The resulting small p value (0.037) indicated that there is good reason to believe that speakers using *-ing* are more likely to be heard as articulate than those using *-in*. Likewise, this test told us that listeners are more likely to select the Midwest when they also select the South. In the latter case, it is important to avoid assuming that one of these causes the other. The data can only tell us the degree to which these two responses are associated, not whether one triggers the other.

To analyze the ratings variables, I used ANOVA. These analyses examined both the influence of the independent variables on ratings and the relationship between

checkbox variables and rating variables. I accomplished this latter task by temporarily treating a checkbox variable as an “independent” variable and using it as a term in an ANOVA run. In a handful of cases, I used a different rating variable as a factor, to investigate its relationship to another rating variable.

There are a number of ways to approach ANOVA, specifically relating to the order in which the terms are considered. This order is relevant because as each term is evaluated, the result indicates the degree of variance in the data which it accounts for, after having controlled for those terms already considered. Some techniques consider all terms simultaneously, while others consider each term after controlling for the effects of all others. A third option is a hierarchical approach and this is the one that I used. In this approach, the order of the independent variables (or those being treated as independent for the purposes of the ANOVA) are considered in an order specified by the analyst. Thus, if the variable indicating which recording is being evaluated is listed first, it will be considered first, then its effects will be subtracted out of the analysis for the next term listed. This approach is useful in cases where the variables being examined can (and should) be ranked by a process which is theoretically prior to running the analysis.

The bulk of the variables used in my analysis of variance fell into a hierarchy. The speaker in the recording or the choice of recording itself clearly accounted for the lion’s share of variance in nearly all the responses. In other words, listeners were most influenced in their judgments by which speaker they heard talking and what the person said. These two variables were the logical first terms. Because they nest within each other, (i.e. different speakers have different recordings) only one or the other may be used in a given run. When considering the data dealing with a single speaker, the term indicating the recording is used as the first (most influential) factor. When considering the data as a whole, however, I used the speaker, rather than the recording. This is because the data includes 32 recordings, which is too large and unstructured a factor to use on its own. Although there is variation within speakers, in the context of a larger dataset, the speaker is the more comprehensible and thus more useful factor.

The next term in every round was (ING). Although its effects were small compared

to the large impact of the rest of the linguistic material, it still had a larger influence than the listener variables. This relates less to the relative importance of an individual linguistic variable and listener variation and more to the fact that only very general information was collected about each listener, allowing only the most basic of analyses regarding their demographic characteristics.

These listener characteristics make up the next set of terms: gender, race, school and regional background. These variables are much more difficult to rank in a theoretically justifiable way. However, since the size of the data set made it difficult to explore interactions between more than one of these in any case, so only one was included in each round of ANOVA. It is certainly possible that interesting interactions between these factors will emerge in later research, but with the current dataset exploring these issues is beyond this study.

The last possible category of terms is the checkbox variables discussed above. Although these variables are, strictly speaking, dependent variables, it is useful to treat them statistically as independent variables in some cases, to explore the relationship between them and the ratings. By doing so, we can learn that listeners rate speakers as more highly *educated* when they select the term *articulate* to describe them. As with co-occurrence patterns between checkbox variables, it is important to avoid assuming that one of these variables causes changes in the other. Because both are dependent variables, it is impossible to determine what causal relationship, if any, exists between them. Also, it is necessary for the checkbox terms to be considered last in the analysis of variants, so as to avoid attributing such a pattern in a case where, the two descriptions are merely correlated by being used to refer to the same speaker. As a final decision, I limited each analysis of variance to three-way interactions, because the number of data points available could not support more complex interactions.

The following chapters describe the results of the study. Chapter 4 goes over the proposed core meanings of (ING) discussed in Chapter 2 and examines how they relate to (ING) in the data here. Chapter 5 takes a close look at the role of region and accent in the data and how perceptions of region shape the effect of (ING) in the speech of different speakers. Chapter 6 explores how differences between listeners and

their social and emotional reactions to speakers play a part in shaping the meaning of variation. Finally, Chapter 7 discusses the the theoretical implications of these findings, with particular discussion of the concepts of social meaning and style.