GENDER STEREOTYPE EFFECTS IN SPEECH PROCESSING

DISSERTATION

Presented in Partial Fulfillment of the Requirements for

the Degree of Doctor of Philosophy in the Graduate

School of The Ohio State University

By

Elizabeth A. Strand, B.A., M.A.

* * * * *

The Ohio State University

2000

Dissertation Committee:

Professor Keith Johnson, Adviser

Professor Mary E. Beckman

Professor Donald Winford

Approved by

Adviser
Linguistics Graduate Program

ABSTRACT

In this dissertation, I investigate the influence of gender stereotypes on low-level speech processing. Based on the notion that stereotypes serve as cognitive shortcuts in cognitive processing (i.e., stereotypical things are processed faster), I test both auditory-only and audiovisual speech perception to determine whether speech produced by speakers with stereotypical voices and faces is processed faster than speech produced by speakers with nonstereotypical voices and faces. My results indicate that there is indeed a processing deficit associated with perceiving speech from nonstereotypical voices and faces, and that gender information does, therefore, influence speech perception.

The first four experiments assessed the levels of stereotypicality of a set of voices and faces, such that a number of faces and voices could be identified for use in stimuli construction for a set of speech perception experiments. We identified two stereotypical voices (male and female), two nonstereotypical voices (male and female), two stereotypical faces (male and female) and two nonstereotypical faces (male and female). To select these voices and faces, we combined measures from multidimensional scaling solutions for voice and face similarity ratings with accuracy and reaction times from speeded voice and face gender classification tasks. Combining results from these two methods let us assess notions about gender stereotypes implicitly, rather than relying on more explicit measures as has been done in the past.

These faces and voices were then used in two speech perception experiments. Participants were presented with words uttered by various speakers, and were asked to repeat the targets as quickly as possible in an auditory naming task paradigm. Response time was measured. In one set of trials, participants were presented with the words only, while in a second set of trials, each word was preceded by a prime in the form of a picture of a face of the supposed speaker of the word. Levels of stereotypicality as well as gender of the faces and voices were mixed and matched such that audiovisual pairings of all possible crosses of face and voice gender, and face and voice stereotypicality, were presented to participants.

Reaction time results indicated that there is a significant effect of voice stereotypicality on reaction times, such that words uttered by speakers with stereotypical voices are faster to be processed and repeated in the speeded auditory naming task. Results from the face-primed auditory naming task replicated the voice stereotypicality effect, and also showed that for female voices, there is a significant priming effect of face stereotypicality on the processing and repetition of the target words. Results also indicated an interaction of face gender and voice stereotypicality in perception of the

face-primed female voices. Interestingly, however, we found a null effect of face priming for the male voices. Our results indicated that at least for some voices, gendered face information significantly facilitates the processing of speech information.

These results support the hypothesis that gender stereotypes do influence speech processing, and that they do it very automatically at a low level of cognition.

To my best friend and partner in life

Frederick Parkinson

who supports me in every endeavor

—and—

To my grandmothers

Evelyn Cook Heaser and Gladys Bergum Strand

who guide me every day

ACKNOWLEDGMENTS

This dissertation would never have come about without the unbelievable amount of assistance and support that I have received over the years from my adviser, Keith Johnson, as well as from my teacher Mary Beckman. To each of them I owe my utmost gratitude and appreciation.

Keith and I arrived at Ohio State at the same time, in the fall of 1993, at which point I was assigned as his research assistant. It was through this early allegiance that I was introduced to the world of speech perception research, and was influenced to alter my own intended path of graduate study in sociolinguistics so that it would be more in line with Keith's work. From those early days until the present, Keith has been nothing but completely supportive of my endeavors as a student and a researcher. He has always been a patient teacher and mentor, and has joyfully encouraged me to incorporate aspects of my other linguistic loves, sociolinguistics and language and gender, into my speech perception work. Keith has been generous to a fault with both his time (of which he has precious little) and his talents (of which he has precious lots), and for this I owe him dearly. Keith is a true force for good!

Mary has also been more helpful to me over the years than I could ever have imagined possible. I thank her for her input into my experimental designs and data analyses, her immensely helpful commentary and guidance on the papers that I have written, as well as the huge amounts of time that she has selflessly given me this summer to help me finish this dissertation on deadline. Her insights have been invaluable to me.

I am also very appreciative of the efforts shown to me by my third committee member and teacher, Don Winford. Even with his hectic research, teaching, and travel schedules, Don has always found time to meet with me, talk with me about my work, and give me immensely helpful feedback.

I must also recognize Norma Mendoza-Denton, whom I like to think of as my language and gender muse. Norma has always been overwhelmingly supportive of my attempts to uncover the links between speech perception and gender issues, and her enthusiasm for this work has been very motivating over the years.

I owe my gratitude to Geneviéve Escure and Andrew "Sandy" MacLeish, as well, who were my first professors of linguistics in the Department of English at the University of Minnesota. Their infectious love of language study and their inclusion of me in the very supportive, intimate collegiate environment that they had created within the much larger English department at Minnesota were the prime factors that influenced me to study linguistics at the graduate level. I loved learning from Geneviéve and Sandy.

The experience of being a graduate student involves more than just interaction with one's teachers. In this vein, I am grateful to the OSU Department of Linguistics for their continuous financial support of me during the past years, as well. Whether I was hired as a teacher or as a researcher, the department's quarterly support has allowed me to pay my bills and keep my refrigerator stocked, while at the same time helping me grow professionally. The department has also always generously funded my travel to conferences and workshops, where I have made crucial professional contacts with other linguists outside of Ohio State. I consider myself extremely lucky to have been able to take advantage of these opportunities. The department has helped me even more this past summer by awarding me a dissertation completion fellowship, allowing me to devote all of my time to my immediate research. Additionally, they have granted me generous funding with which to reimburse my experiment participants. I could not have completed my dissertation this past summer without this support.

I am grateful for several other university-related fellowships that I have been awarded while at Ohio State, as well. I acknowledge the OSU Center for Cognitive Science, the Department of Spanish and Portuguese, and the Department of Linguistics for jointly awarding me a Cognitive Science Summer Graduate Fellowship during the summer of 1997, during which time I was able to explore face perception to a greater extent than had been possible for me earlier. I also acknowledge the Department of Women's Studies for awarding me an Elizabeth D. Gee Grant for Research on Women in the spring of 1998, which helped me to perform the pilot work for this dissertation. These fellowships funded both time and material resources for me, and I am thankful for this assistance.

A good part of the work that I have done during my years at Ohio State, both leading up to and including this dissertation, has been experimental, which necessarily incorporates the efforts of many people. I owe my thanks to all of them. I have made hours of video and audio recordings of faces and voices, from which to create my experimental stimuli. Many graduate students and staff members from this department have volunteered their face and voice talents to this effort, including Ken Bame, Paul Davis, Karin Golde, Jamie Green, Steve Hartman Keiser, Rebecca Herman, Craig Hilts, Stefanie Jannedy, Mike Katkin, Amanda Miller-Ockhuizen, Jen Muller, Frederick Parkinson, Jignesh Patel, Robert Poletto, Jenny Vannest, Neal Whitman, and Steve Winters. Thanks, too, to all of the now-anonymous undergraduate students culled from our linguistics classes who allowed me to record their faces and voices as well.

I also thank Frederick Parkinson and Stefanie Jannedy for the numerous times that they helped me make my video- and audio-recordings when I had other obligations and was unable to do so myself.

Without the assistance of all of these people, I would have no stimuli.

I appreciate the help of the many teachers who have let me visit their linguistics classes over the years to recruit experiment participants, including Mary Beckman, Mike Cahill, Mike Calcagno, Kevin Cohen, Hope Dawson, Janice Fon, Karin Golde, Jamie Green, Toby Gonsalves, Steve Hartman Keiser, Stefanie Jannedy, Hyree Kim, No-Ju Kim, Matt Makashay, Bettina Migge, Nasiombe Mutonyi, Panos Pappas, Frederick Parkinson, Robert Poletto, Craige Roberts, Misun Seo, Tom Stewart, Giorgos Tserdanelis, Jennifer Venditti, Jinyi Wang, Tracey Weldon, and Jae-Hak Yoon.

Graduate and undergraduate students from this department, as well as many friends, have volunteered as experiment participants over the years. I wholeheartedly thank them, as well as the several visiting scholars to our department who have taken part in my experiments as well. These people include Karen Arnold, Catie Berkenfield, Allison Blodgett, Mary Bradshaw, Steve Conley, Sheri Ettefagh, Robin Giampapa, Michelle Gregory, Ed Hubbard, Rich Locke, Laurie Maynell, Amanda Miller-Ockhuizen, Frederick Parkinson, Bill Raymond, Lara Taylor, Jenny Vannest, Patricia Vermillion, and Steve Winters.

I also thank the numerous graduate students from the Department of Speech and Hearing Science, as well as the Department of Psychology, who participated in my experiments.

The bulk of my experiment participants, however, have been undergraduate students from our linguistics classes, and to them I remain most grateful. I hope that their participation in my experiments enhanced their adventures in linguistic study just a little bit.

Without the assistance of all of these people, I would have no data.

My work has been made eminently easier with support from our great linguistics staff members over the years. Jim Harmon and Matt Hyclak have provided me with excellent technical support to meet my every computing need, and often at a moment's notice despite their chaotic work days. Claudia Morettini, Jane Harper, and Maxine Vargas in the linguistics department office have helped me with all types of administrative business, related to both my teaching duties and my status as a student, and I acknowledge them for their efforts.

My fellow student and post-doc officemates in Room 24 of Oxley Hall, a.k.a. "The Labbies," comprise a most cohesive group, and create a very supportive and pleasant atmosphere in which to come to work. I will miss them when I can no longer lay claim to my corner desk in 24B: Allison Blodgett, Janice Fon, Satoko Katagiri, Scott Kiesling, Matt Makashay, Laurie Maynell, Julie McGory, Amanda Miller-Ockhuizen, Giorgos Tserdanelis, Jenny Vannest, Pauline Welby, Steve Winters, Peggy Wong, and Kiyoko Yoneyama.

And finally, I must acknowledge my friends and family, who, best said, are all just a really great bunch of people. I am very thankful for a number of friends who are or have been in similar academic circumstances, and who have been especially encouraging and helpful to me during the time that I have been finishing this dissertation, including Erin Diehm, Svetlana Godjevac, Karin Golde, Steve Hartman Keiser, Craig Hilts, Stefanie Jannedy, Laurie Maynell, Julie McGory, Bettina Migge, Jen Muller, Panos Pappas, Robert Poletto, and Christian Uffmann. I am also extremely lucky to have had two wonderful lifelong friends, Shawn Brommer and Mark Wunderlich, who have been nothing but completely loving throughout this entire process.

My parents Gavin and Constance Strand, as life-long educators themselves, have always placed great emphasis on education and its rewards, and so have been unendingly supportive of my state as a full-time student for most of the past 26 years. I am also thankful for the constant encouragement that I have received from my brother Robert Strand, my sister Jennifer Strand, my brother-in-law Doug Mattson, and my nephew Kai Mattson.

I appreciate the support that my family-in-law has shown me over the years, as well. My parents-in-law Brooke and Shirley Parkinson and my brother-in-law John Parkinson have always been so gracious to me and respectful of my life as a student. My sister- and brother-in-law Sarah and Brian Miller, and my nieces Grace and Carrie Miller have also been unending in the kindness that they have shown me. I appreciate that the entire Parkinson clan has been unquestioning of the fact that "Aunt Liz" chose to take, to borrow the words of Robert Frost, the road less traveled by.

And last but certainly not least, I must thank my dear husband Frederick Parkinson, who has stood by me unyieldingly throughout the ever-chaotic process of finishing this work. While just as busy himself in his own career, Frederick has always made sure that the most important things got done to keep our home life as stable as possible, giving me an always-welcomed oasis of calm—thanks to Frederick, the cat always got fed, the coffee always got made, and the paper always got brought in.

Who could ask for anything more?

VITA

April 12, 1969 .... Born — Rochester, Minnesota

1992 .... B.A. English, University of Minnesota

1995 .... M.A. Linguistics, The Ohio State University

1993 — present .... Graduate Teaching and Research Associate, The Ohio State University

PUBLICATIONS

RESEARCH PUBLICATION

1. Johnson, Keith, Elizabeth A. Strand, & Mariapaola D'Imperio. (1999) Auditory—visual integration of talker gender in vowel perception. *Journal of Phonetics, 27:4,* 359-384.

2. Strand, Elizabeth A. (1999) Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Social Psychology, 18:1,* 86-100.

3. Strand, Elizabeth A. (1999) Gender perception influences speech processing. In Ursula Pasero & Friederike Braun (Eds.), *Wahrnehmung und Herstellung von Geschlecht: Perceiving and Performing Gender* (pp. 127-136). Wiesbaden, Germany: Westdeutscher Verlag.

4. Strand, Elizabeth A., Bettina Migge, Frans Hinskens, & Steve Hartman Keiser. (1998) "TR-/æ/" in the Midlands: A study of the tensing and raising of /æ/ in radio speech in Columbus, Ohio. *Papers in Sociolinguistics: NWAVE-26 à l'Université Laval* (pp. 115-135). Québec, Canada: Éditions Nota bene.

5. Hartman Keiser, Steve, Frans Hinskens, Bettina Migge, & Elizabeth A. Strand. (1997) The Northern Cities Shift in the Heartland? A study of radio speech in Columbus, Ohio. In Kim Ainsworth-Darnell & Mariapaola D'Imperio (Eds.), *Working Papers in Linguistics, Volume 50: Papers from the Linguistics Laboratory* (pp. 41-68). Columbus, Ohio: Department of Linguistics, The Ohio State University.

6. Strand, Elizabeth A. (1996) Gender beliefs and expectations affect fricative perception. In Natasha Warner et al. (Eds.), *Gender & Belief Systems: Proceedings from the Fourth Berkeley Women and Language Conference, April 19-21, 1996* (pp. 403-412). Berkeley, California: BWLG.

7. Strand, Elizabeth A. & Keith Johnson. (1996) Gradient and visual speaker normalization in the perception of fricatives. In Dafydd Gibbon (Ed.), *Natural Language Processing and Speech Technology: Results of the 3rd KONVENS Conference, Bielefeld, October 1996,* (pp. 14-26). Berlin, Germany: Mouton de Gruyter.

FIELDS OF STUDY

Major Field:	Linguistics
Major Focus:	Phonetics
Specializations:	Speech Perception
	Language and Gender
	Sociolinguistics

TABLE OF CONTENTS

Appendices:

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.0 INTRODUCTION

This dissertation is about the role of gender perception in speech perception. Recent experimental results from the speech perception literature suggest that listeners rely on much more than simply the acoustic signal in recognizing speech. Rather, listeners seem to be influenced as well by their social stereotypes about speakers in perceiving the speech signal. This influence implies that even apparently "low-level" or "merely phonetic" constructs that most earlier accounts relegated to an innate "universal phonetics" (Chomsky & Halle 1968) or discounted as aspects of speech that were not "linguistically relevant" (Nolan 1995:4), cannot be predicted completely by the physiology and physics of speech. They are, rather, sociocultural constructs that interact with other sociocultural constructs, including the sex and gender constructs discussed here.

If even "low-level" speaker gender effects are culturally mediated, how can we build models of perception that are grounded in any realistic way in the physical world of articulating vocal tracts and vibrating timpanic membranes? A viable answer to this question would go a long way toward resolving the nature-nurture problem. In this chapter, I first briefly outline one answer that clearly is not viable. This is the "talker normalization" solution, which was adopted in so much of the important classic work on vowel perception in the 1970s and 1980s. Showing how talker normalization does not work sets the stage for considering an alternative approach, which I will present in greater detail in Chapters 2 and 3.

I also take the opportunity in this introduction to discuss the important distinction between the notions of *sex* and *gender*, an understanding of which is crucial to the link that I wish to draw between the perception of social categories and phonetic categories

1.1 Why Talker Normalization Doesn't Work

As was first noted by Aristotle in his *Politics* (328 B.C.), "man [*sic*] is by nature a social animal," and indeed the most socially-encultured animal of all (see Aronson 1995). Human culture and the human mind have evolved to treat all behavior—including the minutest details of speech, gender construction, and even the seemingly most biologically-driven processes such as childbirth (see Mitford 1993)—as inextricably both culture- *and* body-driven. Individual minds come to embody the social constructs of the ambient culture. As "social animals," we cannot divorce ourselves from our social environment, the effects of which filter into even the very lowest levels of cognition, including speech perception.

Interestingly, even the research area of psychiatry has begun to realize the importance of acknowledging the close links between human culture and the human mind:

> [In reviewing] the development of concepts about the contribution of nature and nurture to brain structure and mental function, . . . contemporary psychiatric research conclusively demonstrates that the mind/brain responds to biological and social vectors *and is jointly constructed by both*. Major brain pathways are specified in the genome; detailed connections are fashioned by, and consequently reflect, socially mediated experience in the world. [Eisenberg 1995:1563, emphasis added]

While perception researchers have accepted for some time that listeners use extra-acoustic, visual information about the formation of speakers' lips, jaw, and visible vocal tract during speech perception (as in the McGurk effect and speechreading processes; see Chapter 3), the phenomena that are of interest here indicate that perceived speaker characteristics such as sex/gender and nationality also affect the perception of speech. For example, Strand and Johnson (1996) showed that merely changing the gender of the face in an audiovisual speech token was enough to induce a reliable shift in perception of the acoustic speech signal.

I argue that such perceptual phenomena result from the influence of socially-motivated stereotypes, related to these speaker characteristics, during speech processing. I believe that incorporating recent findings from the disciplines of social psychology, stereotype and face perception research, and gender research into our speech perception work will help us to better account for these socially-mediated phenomena.

While consideration of the impact of social stereotypes as such on speech perception is a recent undertaking, the incorporation of social information into linguistic study in general is not a new occurrence. As Niedzielski (1997) discussed, researchers in certain linguistic subfields such as sociolinguistics and the related field of sociology of language have long considered the interaction between social information about speakers and the evaluation of their language usage.

Lambert and colleagues in 1960, for example, introduced their "matched guise" technique for assessing the varying social interpretations that accompany speakers using different guises of language. They examined the influence of language variety on social evaluation of the speaker, concluding that "language variety alone can affect how the speaker is perceived" (Niedzielski 1997:10).

Thakerar and Giles (1981), on the other hand, studied the ways in which social information about the speaker influences the evaluation of the language used. When they told listeners that a speaker was highly intelligent, for example, the listeners perceived a significantly faster speech rate and more standard accent than when the researchers suggested that a speaker was less intelligent.

But even prior work in the field of speech perception-proper hints at the influence that social information drawn from the voice might have on processing the speech signal. Ladefoged and Broadbent (1957), for example, conducted some of the first perceptual work to suggest that seemingly extra-phonetic information in the voice actually has an effect on phonetic categorization. Ladefoged and Broadbent (1957) showed that perception of vowels in monosyllabic test words was affected by altering the formant structure (the vocal tract resonance frequencies) of synthesized vowels in carrier sentences. Their results indicated that the personal information about physiological and anatomical shape of the vocal tract (information which naturally varies from talker to talker and is inherent in all of the vowels spoken by a given talker) influences the listener in determining the final vowel percept.

More recent examples of perception research that imply socially-linked influences include Johnson (1990), which examined the effects of fundamental frequency (F0) variation in a carrier phrase on vowel identification. Johnson concluded that "contextual talker-related information," or assumptions about speaker identity, influence phonetic identification. Mullennix, Johnson, Topcu-Durgun, and Farnsworth (1995) found that voice information about gender is retained in memory and used in phonetic discrimination; it is not simply normalized away prior to phonetic categorization, as many researchers had assumed was the case. Mullennix and colleagues (1995:3091) suggested that these rich representations are "probably an auditory composite of the various acoustic factors relevant to voice gender like F0, formant frequencies, breathiness, etc."

We can contrast the approach that is emerging from this work with the approach that was dominant in the 1970s and 1980s—namely, *speaker normalization*. For some time, speech perception researchers have wrestled with how to explain the human perceiver's ability to process infinitely varying acoustic speech signals from an infinite number of speakers, and to recover (or uncover) those speakers' intended linguistic messages. Additionally, the complex speech signal is dense with a number of different kinds of highly variable information that must somehow be correctly used by the perceptual system in proper relation to specific speakers, in order to render speech recognizable.

Traunmüller (1994) offered one description of the many sources of variability in the speech signal. His four-part taxonomy of the basic types of information found in the speech signal includes what he termed phonetic information, affective information, personal information, and transmittal information. He defined *phonetic information* as encompassing the "pure linguistic information" in an utterance, which is "reflected in [an utterance's] 'phonetic quality' and can be reproduced in a narrow phonetic transcription of the usual kind" (p. 172). *Affective information* was defined as the "paralinguistic" information that also contributes to an utterance, including, for example, speaker attitude as reflected by vocal effort, pitch range, and intensity, and expression of emotions. *Personal information* was characterized as "all of the extralinguistic information . . . [that is] merely informative about [a speaker's] age, sex, physiological state (pathology), psychological state (unintentional effects of emotion) and environmental state (e.g. the Lombard effect)" (p. 173).[1]

Traunmüller pointed out, however, that these "extralinguistic" speaker states "can, to some extent, be faked." Moreover, since these states so often contribute to what he classified as affective information, it is, in effect, often difficult to tease apart personal and affective qualities, an important point to remember for the purposes of this work. It is also crucial to remember here that, following our recent findings that information about gender actually affects low-level speech perception, it is no longer wholly accurate to ignore such information in linguistic accounts of such processing on the grounds that it is "extralinguistic."

The fourth type of evidence identified as inherent in the speech signal, termed *transmittal information* by Traunmüller, is information about the speaker's location and the specific environment's acoustics, whether acoustic transfer is direct to the ear or over a telephone line, and so on. This information may seem to contrast with the personal information in being somewhat less easy to fake, since it is information about the acoustic channel rather than about the source. However, effects such as the skilled illusions about location that a ventriloquist can create show that transmittal information also is guided by learned expectations.

Even such a basic description of the various information present in the speech signal makes it clear that recognizing the *speech* within the complex signal is no simple task. As all speakers' vocal tracts, idiolects, and other personal characteristics affecting language usage differ, so too will the speech signals that individual speakers produce. Even within a single speaker, great variation in general acoustic output can be observed with changes in, e.g., context and topic of speech, addressee, and emotional state.

Moreover, all of these "extralinguistic" characteristics can vary dynamically over the time course of a discourse. For example, a speaker's vocal tract length, which may seem at first to be a relatively simple invariant "personal characteristic," changes constantly during speech production, as I will elaborate on at a later point.

---

[1] The Lombard effect refers to the phenomenon that people tend to speak louder in the presence of background noise.

As I mentioned briefly above, traditional accounts of speech perception have simply assumed that these speaker-specific (or "indexical") features of speech constitute little more than "noise" in the speech signal and are something to be normalized away to reveal the "true" linguistic information wrapped up in the complex signal. In explaining speech perception, these traditional accounts posit that chunks of the acoustic speech signal are transformed in such a way that they can then be identified as belonging to abstract linguistic categories (or "idealized symbolic representations," as Pisoni 1997 refered to them). It is through this transformation process that indexical information is assumed to be normalized away, rendering a signal that is reduced for speaker differences.

Acoustic normalization models, such as those described in Bladon, Henton, and Pickering (1984), Miller (1989), Syrdal and Gopal (1986), and Traunmüller (1981), use various methods to reduce the differences among speakers before linguistic categorization occurs. Bladon et al. (1984), for example, employed a method of sliding auditory spectra one Bark unit either up or down the Bark scale before comparison with vowel templates, depending on whether the speakers were judged to be male or female. Miller (1989) used logs of the formant frequency ratios with F0 averaged across time to "maintain effective elimination of the effects of talker age and sex" (Miller 1989:2130). Both Syrdal and Gopal (1986) and Traunmüller (1981) compared the values of specified formant frequencies, the differences of which were then essentially matched to "templates representing phonetically significant patterns" (Traunmüller 1981:1472).

But the problem with perception models such as these is that they grossly oversimplify the impact that indexical information actually has on phonetic categorization. Indexical information in the speech signal is not simply noise to be "eliminated," but rather significantly informs the phonetic categorization process, as the results reviewed by Mullennix et al. (1995), for example, suggest. Models such as these fail to take into account that although we can make gross generalizations in comparing across speakers or speaker groups, there is also a good deal of variation within the speech of a single speaker, reflecting both "phonetic information" (i.e., socially-constructed symbols or indices related to lexical categories and the like) and "affective" and "personal" information (i.e., socially constructed indices related to other behavioral categories such as gender).

Furthermore, this variation is under speaker control to a far greater extent than the normalization literature has ever acknowledged. Fundamental frequency, for example, is not simply a static artifact of the size and shape of the speaker's thyroid cartilage, but is rather a dynamic property, a socially-linked behavioral variable which reflects and is used for a number of purposes—including the cuing of intonational category *and* the culturally-specific cuing of a speaker's desired gender category. "Normalization" by F0, as in Miller's (1989) simplified treatment, neither recognizes, predicts, nor accounts for these sources of F0 variability.

In short, a good deal of current perception research suggests (contra the aforementioned theories that rely on speaker normalization) "that the processing of voice and phonetic information are closely tied together" (Mullennix et al. 1995:3080). This recognition of the failure of the classic "normalization" model has opened the way to a new treatment of "personal characteristics." Stereotypes, like those studied for years in sociolinguistics and psychology of language, have come to be considered factors to be accounted for in speech perception in ways other than being information that is simply normalized away.

In Strand and Johnson (1996) and Johnson, Strand, and D'Imperio (1999), for example (both of which will be more fully described in Chapter 3), we examined the influence of gender stereotypes on the perception of fricatives and vowels. In work such as this, what would earlier have been conceptualized in terms of automatic adjustments made through normalization transforms, is beginning to be understood, rather, in terms of potentially highly culturally-specific learned behavior.

Our work involved audiovisual tokens—short movies of faces saying various words. Our fricative study examined perception of a synthesized frequency continuum between the fricative speech sounds /ʃ/ and /s/, as in the words *shod* and *sod*. In this study, we found that we could induce a shift in categorization of the fricative portion of the auditory signal merely by changing the gender of the face shown in the visual signal. Participants perceived the auditory signal differently depending on whether they saw a female face or a male face "producing" the signal.

In the case of our fricative continuum, participants were more likely to perceive the auditory portion as *sod* when they saw a male face producing it, and more likely to perceive the auditory portion as *shod* when they saw a female face producing it. That is, in male-faced tokens, the boundary between fricative categories was perceived to be at a lower frequency, with a greater number of experimental tokens therefore classified as *sod* (with a higher-frequency fricative onset); in female-faced tokens, on the other hand, the boundary between fricative categories was perceived to be at a higher frequency, with a greater number of experimental tokens therefore classified as *shod* (with a lower-frequency fricative onset). These results led us to suggest that expectations for how male and female speakers "should" sound (expectations which could be activated just by information from the faces in the audiovisual tokens) affected the perception of the fricatives.

Findings from our vowel work were similar. In these studies, we examined the perception of a synthesized F1 frequency continuum between the vowel sounds /ʊ/ and /ʌ/, as in the words *hood* and *hud*. Again, we found that by manipulating the gender of the face in the audiovisual token, we could induce a shift in the phonetic categorization of the vowel. When the face in the audiovisual token was male, listeners were more likely to perceive the auditory portion as *hud* (with the F1 frequency boundary at a lower frequency), and when the face was female, they were more likely to perceive the auditory portion as *hood* (with the F1 frequency boundary at a higher frequency). Again, we concluded that gendered aspects of the face were invoking certain expectations for how speech coming from a male or a female "should" sound, since the auditory portions of the audiovisual tokens were held constant.

Interestingly, we found that even the mere *suggestion* that a speaker was male or female (in an audio-only perception test) was enough to induce a reliable category shift in identification of the vowels. When we told listeners that an auditorily androgynous speaker was male, they perceived more of the vowel tokens as *hud.* When we told the listeners that the same speaker was female, they perceived more of the same vowel tokens as *hood.* This is strong evidence that gendered expectations about speakers, even expectations raised merely by suggestion, influence low-level phonetic categorization.

In a similar audio-only vowel categorization task, Niedzielski (1997, 1999) showed that the suggestion that a speaker was American or Canadian influenced vowel classification in much the same way. She therefore concluded that social expectations based on assumed nationality also affect phonetic categorization (see Chapter 3 for a more complete description of Niedzielski 1997, 1999).

So, we want to know: how is information about gender handled in speech perception? Obviously, the voices of men and women, and boys and girls, are in general acoustically quite different from each other, and tend to be perceived as quite different from each other, a notion supported by the generally high levels of accuracy reported in listeners' discrimination between females and males.[2] Indeed, babies as young as 7 months old have been shown to accurately discriminate between male and female voices (Walker-Andrews, Bahrick, Raglioni, & Diaz 1991:57 cited Miller, Younger, & Morse 1982).

---

[2] Mary Beckman (personal communication) suggests that speakers could be less accurate at identifying sex for unfamiliar cultures with rather different gender-voice practices. Unfortunately, I know of no research that has addressed such cross-cultural gender-via-voice identification. If this suggestion could be borne-out empirically, however, it would constitute strong evidence that such knowledge is learned (i.e., cultural), and not simply noise that can be normalized out of the signal by an algorithm that is insensitive to cultural background.

One standard explanation offered by speech perception researchers has been that listeners simply normalize the signal to the perceived vocal tract size of speakers, using any acoustic (or visual) information available that may help them to accomplish this task. As we reviewed in Johnson et al. (1999), this information could include both direct cues (aspects of the acoustic signal that give a direct indication of vocal tract length, such as formant frequencies) and indirect cues (aspects of the signal that are not directly contingent on tract length but can be correlated with it, including F0). I will call such an account, in which the listener forms a perceptual estimate of the physical characteristics of the speaker, a "vocal tract normalization" theory.

So, how might the vocal tract normalization theory account for the phenomenon that we identified in Strand and Johnson (1996), where changing the gender of the face of a talker influenced the phonetic categorization of the speech signal? It could be proposed that at least a cursory estimation of vocal tract length could be assessed from seeing the head and neck of the speaker (i.e., somehow calculating the vocal tract length from the visual display), information which could then be used in the normalization process. In general, the vocal tract of a voice that is presumed to be male because of the simultaneous visual display of a male face, would likely be assessed as longer than that of a voice presumed to be female.

But this proposition, as we pointed out in Johnson et al. (1999), is naïve in a number of ways, principally because the vocal tract is not a static entity, but rather is constantly changing in length during speech production. It might typically be longer in [d] than in [t] and shorter in H* than in L*, other things being equal. But these typical differences in larynx height are something that cannot be directly assessed from the visual display of a speaker's face. The entire normalization approach cannot accommodate the dynamic state of the vocal tract, and it clearly offers only a gross oversimplification of the observable variability in the speech signal.

Additionally, an explanation that relies on the calculation of vocal tract length cannot account for the suggestion effect that was described above, where we influenced listeners' phonetic categorization of the speech signal not by showing a picture that gives visual cues to vocal tract length, such as relative head size, but merely by *telling* them that the speaker was a female or a male. Niedzielski's (1997, 1999) work on the effect of suggested nationality of the speaker on phoneme categorization is an additional example of a perceptual influence that cannot be addressed by a traditional vocal tract normalization approach. I hypothesize that these suggestions of speaker gender and nationality invoke expectations, or stereotypes, in the minds of the listeners, which are somehow incorporated into their perceptual processes and impact the final phonetic categorizations.

Stereotypes have been considered to have an effect on many cognitive processes, including the presumably low-level perception of speech rate as described earlier in Thakerar and Giles (1981), where the investigators found that a stereotype about intelligence affected perceived rate. Results such as these, Niedzielski's, and ours force speech perception researchers to reconsider the ways in which they formulate the perception process. How can such low-level stereotype effects be explained and incorporated into speech perception theory, and how can we determine how truly "low" stereotype effects go in cognitive processing? To begin to answer these questions, it is necessary to generally understand how stereotype research to this point comes into play with the speech perception results that have been briefly discussed here.

For nearly a century now, researchers in the field of social psychology have examined the ways in which social stereotypes about groups of people affect the way those groups are generally perceived by others. A good deal of attention has been paid to assessing the ramifications of individuals' association of specific personality characteristics with other people; such associations have been shown to affect other cognitive processes (including further stereotype formation, maintenance, and application) in a number of significant ways. Of specific interest to our audiovisual speech perception studies are the ways that individuals' physical attributes have been shown to induce the perception of distinct personality traits, which are then extrapolated to general stereotypes about the individuals. This link has long been the subject of popular and even pseudo-scientific interest. From Cicero's declaration in ancient Rome that "the countenance is the reflection of the soul" (quoted by Zebrowitz 1997:2), to the road-side fair phrenologists common into the early twentieth century, people have believed that physical appearance is closely linked to personality. These links lead to stereotypes that people then use to interpret further behavior from the stereotyped individuals.

Interesting effects such as the "babyface effect," which I will further describe in Chapter 2, offer evidence in support of the notion that people really do make stereotyped judgments about others based on the way they appear. The babyface effect was first reported in the early 1980s (McArthur & Apatow 1983-84), and implied that for adult faces, "larger eyes, shorter features, and lower vertical placement of features, which yields a larger forehead and a smaller chin, each increased perceiver's impressions of a target's weakness, submissiveness, and naïvete" (Zebrowitz-McArthur & Montepare 1989:190).

Impressions of such character attributes as linked to facial traits, and the resulting social stereotypes about individuals with those traits, seem to be consistent across perceivers within cultures. Similar research has been conducted to assess attractiveness and linked attributes. The general conclusion of this research is that "attractive people of both sexes and all ages are perceived to have more positive traits" (Zebrowitz 1997:158). The halo effect[3] then influences perceivers to assess other aspects of attractive people's personalities and behavior more positively than would be the case for less-attractive people.

---

[3] The halo effect is defined as implying that "a liked person [or in this case, an attractive person] is assumed to have good qualities of many kinds, whether or not the observer has any information about those qualities" (Taylor, Peplau, & Sears 1994:556).

In sum, these observations suggest that appearance invokes stereotypes, and these stereotypes affect other aspects of cognitive processing. The speech perception studies briefly described above give evidence that stereotypes also affect the processing of language.

As I am interested in examining the interaction of gender stereotype categories with phonetic categories in speech processing, it is necessary for me to clarify what I mean by *gender stereotype categories.* Therefore, before addressing the more linguistic aspects of this dissertation in Chapters 3, 4, 5, and 6, as well as the notion of stereotypes in Chapter 2, I will first define the ways in which I will use the terms *sex* and *gender* here, as well as place their currently-understood relationship into its historical context.

1.2 *Sex* vs. *Gender:* Review and Clarification

The terms *sex* and *gender* reflect two very complex notions, and attempts to define and differentiate the two have received much recent attention in both the popular and scholarly presses. While the history of the quest to biologically characterize female and male differentiation takes a long and interesting path, today's commonly-accepted (Western) notion of *sex* as a binary category based on biological divergence came to the fore at the end of the eighteenth century. (See Laqueur 1990 for a treatment of *sex* throughout history, including discussion of the politically-motivated transformation from assuming a one-body, sex-continuum view to our "modern" two-body, two-sex view.)

Stemming from the notion of sex as a binary categorization based on anatomical and physiological differences between males and females come notions of socially-appropriate and expected behavior, e.g., men are thought to be (and expected to be) more assertive and independent than women, and women are likewise assumed and expected to be more emotional and concerned with others' welfare (Deaux & LaFrance 1998:793, citing Deaux

& Lewis 1984; Spence, Helmreich, & Stapp 1975). Most, if not all, aspects of behavior can be seen as similarly organized along male-female lines, including such obvious features as appropriately-deemed make-up and jewelry usage, hairstyle, and dress.[4]

With the desire to emphasize that such behaviors are socially motivated (and so are, therefore, arbitrarily linked to females and males, rather than being intrinsically biologically determined), feminist psychologists in the late 1960s borrowed the term *gender* from the field of linguistics. Linguists had long used this term in reference to the arbitrary assignment of nouns, pronouns, and modifiers to classes, or "genders," which were then inflected accordingly as specified by the language. Feminist psychologists found the notion of arbitrary category assignment in linguistic gender to be a useful analogy in their own research aimed at highlighting and describing the socially-prescribed behaviors that for so long had been considered simply direct consequences of being male or female. Unger and Crawford (1993:123) summarized the dual-usage of the term *gender* as such:

> In grammar, gender is understood as a way of classifying phenomena—a socially agreed upon system of distinctions rather than an objective description of inherent traits . . . [in a similar way], societies prescribe particular characteristics for males and females on the basis of assigned sex (Vaughter, 1976; Weisstein, 1968). Thus, gender must be examined as a cultural as well as a linguistic phenomenon.

Early researchers in the emerging field of gender studies based their notions of gender on existing models of personality, which held that males were characterized by a set of traits considered "masculine," and females were characterized by a set of traits considered "feminine" (Deaux & LaFrance 1998:799, citing Terman & Miles 1936). These models assumed that the masculine and feminine traits are arranged on a unidimensional

---

[4] It is an interesting sidenote to consider the particular social institutions that continue to reinforce such expectations: as recently as last year, a female student preparing to graduate from an Albany, Ohio, high school was forced to petition the school board for permission to wear pants beneath her graduation robe during her high school commencement, rather than a skirt or dress as mandated by school policy. Only after seeking legal advice did the board grant the girl's request (Glenn 1999).

continuum, with "most masculine" traits at one extreme and "most feminine" traits at another (Deaux & LaFrance 1998:799, citing Constantinople 1973). According to this view, "one can be masculine or feminine, but not both" (Deaux & LaFrance 1998:799).

But this early simplistic view of gender seemed problematic to many feminist psychologists who recognized that individuals could actually possess stereotypically "feminine" characteristics as well as "masculine" characteristics. "Hence a person can be simultaneously 'aggressive' (judged to be a masculine trait) and 'affectionate' (seen to be more socially desirable in a woman)" (Graddol & Swann 1989:33). The introduction of the notion of *androgyny* to the field in 1974 (Bem 1974) addressed this point, maintaining that healthy individuals could possess traits that are "considered stereotypically appropriate for both sexes" (Unger & Crawford 1993:123) and that "masculinity and femininity are separate and orthogonal dimensions. Thus, people could be masculine, feminine, or androgynous, the latter combining both domains" (Deaux & LaFrance 1998:799).

More recent research on gender theory, driven by social psychological approaches, stresses the importance of considering the context in which gendered behavior occurs. That is, gender is not just a set of static characteristics that exist immune from the surrounding social environment, but rather, as Deaux and LaFrance (1998:788) put it:

> [Gender is] a dynamic construct that draws on and impinges upon processes at the individual, interactional, group, institutional, and cultural levels. As such, the presented analysis is of gender-in-situated-action . . . Social structures, social roles, power, status, and culture are factors and levels of analysis that must be considered for a fully-drawn picture of gender.

In a related way, feminist scholars such as philosopher Judith Butler (1990) propose that gender is *not* a "monolithic construct" (as Cameron 1997:62 put it) that is achieved in completed form at some point during one's existence, thereafter remaining an unchanging bundle of personal characteristics. Gender, rather, is a dynamic ongoing "performance" (Butler's term) that an individual *does* in interplay with a multitude of cultural factors that make up his or her reality. Cameron (1997:49) summarized Butler's performative view of gender as:

> 'constituting the identity it is purported to be' . . . [Just as] illocutions like 'I promise' do not describe a pre-existing state of affairs but actually bring one into being, so Butler claims that 'feminine' and 'masculine' are not what we are, nor traits that we *have*, but effects we produce by way of particular things we *do* . . . This extends the traditional feminist account whereby gender is socially constructed rather than 'natural', famously expressed in Simone de Beauvoir's dictum that 'one is not born, but rather becomes a woman'. Butler is saying that 'becoming a woman' (or a man) is not something you accomplish once and for all at an early stage of life. Gender has constantly to be reaffirmed and publicly displayed by repeatedly performing particular acts in accordance with the cultural norms (themselves historically and socially constructed, and consequently variable) which define 'masculinity' and 'femininity'. [original author's emphasis]

So, in a Butlerian view of gender as a continuously changing and evolving "performance" that we do, we constantly reformulate and redefine our own gender identities as part of functioning in and reacting to the continuously changing contexts that make up our lives. Thus, I perform my own gender in very different ways (a performance that is continuously altered throughout my lifetime) when I am playing different roles and interacting with different people in varying contexts; my performance varies when I am teaching in front of my class, versus vying for service at a busy shop counter, versus riding on a crowded bus to work, versus eating dinner in the backyard with my husband, versus talking with my sister and other female friends, etc.

It is worth emphasizing that social psychologists have long considered the notion of context to be a most crucial factor in predicting behavior. For example, Fiske and Taylor (1991:5) offered the simplified example that "to know that a particular person is motivated to study does not predict whether or how much he or she will study. But a motivated person in a library [the context] is extremely likely to study a lot . . . social psychologists [see] both the person and the situation as essential to predicting behavior."

Similarly, language and gender researchers have recently emphasized the important role that context plays in constructing gendered identity, which underlies our gendered language use, by embracing the concept of *communities of practice* in their research and description of "the orderly heterogeneity of language in its social setting" (Holmes & Meyerhoff 1999:173). The importance of context as well as a Butlerian notion of the continuous construction and evolution of gendered identity come together in the notion of communities of practice, defined by Eckert and McConnell-Ginet (1999:186) as:

> [aggregates] of people who, united by a common enterprise [the context], develop and share ways of doing things, ways of talking, beliefs, and values—in short, practices . . . The development of shared practices emerges as the participants make meaning of their joint enterprise, and of themselves in relation to this enterprise [the continuous development of self-identity] . . . The community as a whole constructs a joint sense of itself through the relation between its practices and those of other communities. Thus a community of practice is not isolated and inward-looking, but shapes its participants' relations both among themselves and with the rest of the world.

(See the entire June 1999 issue of *Language in Society*, edited by Janet Holmes, for an examination of the most current work on communities of practice in language and gender research.)

While all of this may seem to contribute to an overly pedantic notion of the psychology of gender, it is important to note that state-of-the-art research on gender is repeatedly supporting the concept as a very complex one indeed, much more intricate than earlier explanations of the single masculine-feminine continuum could have predicted. (See Beall & Sternberg 1993 for a general overview of the kinds of gender theory research that are currently being undertaken.)

But putting discussions of gender theory aside for the moment and returning to the main point of this section (which is to define *sex* and *gender* as these terms will be used hereafter in this work), it seems reasonable, following my previous points, to generally describe *sex* as being the dichotomous classification of humans into the categories "female" and "male," based on biological indicators, and *gender* as being the related socially-determined qualities and expected behaviors that come to be associated with males and females in a given culture.

While this seems like a very reasonable way to qualify the notions of *sex* and *gender*, unfortunately it is not this simple. Myriad gender researchers throughout the past decade have pointed out that this very reductive notion of *sex* in particular is not wholly accurate. While the notion of sexual differentiation has tended to be viewed as innate since its historical emergence (see Laqueur 1990), many current researchers argue that indeed "*both* sex and gender are socially constructed" (Bing & Bergvall 1996:7, emphasis added), which greatly complicates our previously rather clear-cut notion of sex.

The argument for sex as a social construct draws support from examinations of the treatment of children born with essentially ambiguous sex characteristics, as well as the various cultures throughout the world which foster sex classification beyond a standard bipartite model. Beginning with the evidence for socially-constructed sex from inconclusive biology, Bing and Bergvall (1996:8) pointed out that:

> . . . there is no standard legal or medical definition of sex. Biological sex results from variations in chromosome combinations, internal gonad structure, external gonad structure, hormonal dominance, secondary sexual characteristics, apparent sex, psychological sex, and sex of rearing. In the majority of human births, the combinations of these factors lead to clearly sexed males and females, but they can also result in as many as seventy different types of intersexed individuals (J. Epstein 1990:105). Such intersexed individuals are not as rare as most people think (Duckett & Baskin 1993).

In our culture, oftentimes the first question asked upon hearing of the birth of a new baby is, "is it a girl or a boy?" Consequently, the "failure" of an intersexed individual to fit easily into one of the two established sex categories quickly becomes interpreted as a situation in need of fixing by the medical community. Social pressures for a child to be clearly labeled a "girl" or a "boy" influence physicians to essentially reconstruct biology to conform to expectations of what sex "should be."[5] Continuing from Bing and Bergvall (1996:9):

> . . . intersexed individuals (who were previously treated as monsters) are now defined as 'treatable,' with physicians reconstructing the body as either female or male with surgery and/or hormones. The possibility of *not* 'curing' these individuals is never considered. [original authors' emphasis]

As Glick and Fiske (1999:365) pointed out, it is interesting to consider why it is that our culture is so intolerant of ambiguity regarding a person's sex category, especially in light of the fact that "there is considerable evidence that sex is neither simply dichotomous nor necessarily internally consistent in most species" (Unger & Crawford 1993:24) or in many cultures (Bing & Bergvall 1996:10 cited, e.g., the *hijras* of India, as treated in Hall & O'Donovan 1996; and the *nadle* of Navajo culture, as treated in Martin & Voorhies 1975). Concluding from Bing and Bergvall (1996:3):

---

[5] The following anecdote from a popular television show, described in Glick and Fiske (1999:365), dramatically illustrates this point: "The scene is a birthing room, with the mother in the late stages of delivery, the excited father recording it all with his video camera, the doctor ready to receive the newborn infant. 'You have a beautiful baby . . .' The doctor's voice trails off. 'A beautiful baby what?' demands the father. 'I don't know,' the doctor responds. 'I can't tell.' So begins an episode of the television hospital drama *ER*. Subsequent tests reveal that the baby is chromosomally male; however, it would be easier medically to perform surgery to give him female genitalia (more extensive surgery would be needed for male genitalia, which would not be 'fully functional' in adulthood). With simple surgery and hormones, he could be a 'normal' she (though unable to reproduce). The discussion of how to proceed is dominated by the father, who had counted on having a son to carry on the family name. Because he does not want a son who will be impotent or a daughter whose gender appearance must be maintained through hormones, he decides that he and his wife should give the infant up for adoption."

> Recently, Butler (1990, 1993), J. Epstein (1990), Bem (1993), Nicholson (1994) and others have claimed that, like gender, sex is socially constructed and better described as a continuum rather than a dichotomy. In what are often called Western (meaning industrialized) countries, biological sexual variance is now surgically corrected to fit binary categories (J. Epstein 1990), in contrast to cultures that recognize more than two biological categories, as well as more than two social and linguistic categories (Jacobs & Cromwell 1992, Hall & O'Donovan this volume) . . . Butler, Bem, Nicholson and J. Epstein do not assume dichotomies in either sex or gender and their work is encouraging those in other disciplines to examine the consequences of looking for and finding dichotomies.

Researchers interested in revising the commonly-assumed notion of sex as a natural dichotomy to a notion, instead, of socially-constructed oppositions are driven by the desire to gain a more accurate understanding of human behavior. Traditional sex categories and assumptions about people derived from their membership in one or the other category drive the tendency to stereotype and over-emphasize homogeneity within categories and difference across categories, which thereby serves to erase much of the diverse behavior evident from individual to individual. More recent approaches to understanding sex categorization (as well as gender) seek to avoid the oversimplification that exists in so much existing discussion of differences between currently-classified sex groups.

Despite the fact that females and males have much more in common biologically and behaviorally than they differ in these aspects (see, e.g., Fausto-Sterling 1992:25-59), most sex-related research outside the realm of modern feminist and social psychological theory continues to embrace the bipartite sex model. The reasons for this are many. Speech perception researchers, for example, have historically had little influential contact with feminist gender theoreticians, so it is not surprising that these researchers continue to employ a relatively simplistic understanding of the "categories" female and male in their exploration of speech perception processes.

Furthermore, social psychologists emphasize that in terms of cognitive processing in general, there exists a strong cognitive impetus to employ previously-established categories in classification (an issue that I address further in Chapter 2), which makes nontraditional notions of gender categories less likely to be operative within contemporary humans' cognitive systems. Nonetheless, it is still worth considering the acceptance of a theory of sex/gender that goes beyond the inadequate bipartite sex model for the purposes of more accurately describing and predicting human behavior.

Which brings me back to the original intention of discussing how the concepts of sex and gender will be used in this work. As the theory behind a socially-mediated notion of sex is still relatively difficult to employ in speech perception for the reasons stated above, I will continue to define *sex* as was done previously—the dichotomous classification of humans into the categories "female" and "male," based on biological indicators—but will add the stipulation that this notion is subject to revision as informed by subsequent developments in gender theory research. Following this, I will continue to describe the notion of *gender* as the socially-determined qualities and expected behaviors that come to be associated with males and females in a given culture.

But my intent here is to go beyond a characterization of male and female speakers as simply vocal tracts differentiated by physiology alone, to one that considers speakers as social beings who exist in a social context which influences all aspects of their personal performance, including the way they shape and use their voices, as well as perceive others' voices. For that reason, in addition to the fact that the stereotypes I will be discussing are socially mediated, I will use the term *gender effect* (rather than *sex effect*) to refer to the nature of the stereotype influences that we have shown to occur in speech perception.

1.3 Overview

On the following pages, I will show how gender stereotypes affect phonological processing all the way down to even the very lowest levels of cognitive speech processing. To better understand this, I have organized this dissertation in the following way. In Chapter 2, I present necessary background information on stereotype theory and gender stereotypes. In Chapter 3, I describe a number of previously-reported stereotype effects from several speech perception experiments. These results lead us to understand speaker normalization and all of phonetic processing as being couched in a framework of "stereotypes as cognitive shortcuts," which paves the way for the gender stereotype perception results that I discuss in Chapter 4. In Chapter 5, I present a set of speech perception results that build upon the gender stereotype perception results described in the previous chapter. I conclude my examination of the role of gender perception in speech perception in Chapter 6.

CHAPTER 2

STEREOTYPE THEORY AND GENDER STEREOTYPES

2.0 INTRODUCTION

"It's like the story of the man in the woods," Don Winford said to me one recent afternoon. "You know the one, where the man shoots another man, but afterward he *swears* that he thought he was shooting at a deer. How is it that he can perceive something that's so different from what's really there?"

The question of how the human sensory–perceptual system might come to interpret a particular stimulus as something different from a simple aggregate of the qualities that are inherent to that stimulus is a complex question indeed, one that could keep a bevy of psychologists, neurologists, cognitive scientists, and other interested parties engaged for years. But part of the answer to the question about the man in the woods lies with the expectations that the man holds, based on the context of his perceptual experience. If the man expects to encounter a deer moving through the woods, he is much more likely to interpret any sounds or movement that might seem to be emanating from a similarly-sized animated object, as coming from a deer. His expectations influence his perception.

In a similar way, our expectations about other people, based on the way we perceive their membership in certain social categories, influence our perception of all of their behaviors, from their reaching motions and gait to their smiles and phoneme segments. These socially-mediated expectations can also be thought of as stereotypes. In this dissertation, I will adopt the technical term *stereotype* and the understanding of what stereotypes do from recent social psychology research. This notion is closely related to (and obviously stems from) the lay notion of "stereotype" that we see in the popular press accounts of race- or gender-based prejudice.[1]

However, in the course of investigating stereotypes over the past 30 years, social psychologists have developed a more elaborate cognition-based model that is relevant for my research. As I will describe later in this chapter, this model assumes that perceivers are cognitive misers who use all possible means to help them simplify the cognitive processing of incoming information. Stereotypes are one means to help perceivers in the simplification of this processing.

In the following sections, I describe research on stereotypes as it is relevant to the focus of this dissertation. First, in §2.1, I address the influences that work in the related area of social psychology of language has had on the present research. In §2.2, I give a brief description of the ways that stereotypes have been conceptualized to this point, before I define stereotypes in §2.3 and discuss some of the basic theories about the representation, formation, maintenance, and application of stereotypes in cognition in §2.4 and §2.5. In §2.6, I offer a summary of the large body of research that has addressed the ways that people's faces invoke stereotypes about their personal attributes. This work is relevant for explaining the results of our past audiovisual speech perception experiments, which I discuss in Chapter 3, as well as for motivating the experimental work that I describe in Chapters 4 and 5. Finally, in §2.7, I present a discussion of the current research on gender stereotyping, which is generally relevant in describing our face gender and gender suggestion effects.

---

[1] While there are, of course, stereotypes based on most every kind of human characteristic, race- and gender-based stereotypes are frequently noted as the most ubiquitous in industrialized societies. See, e.g., Chideya (1999), Funderburg (1994), and Sowell (1981) for popular treatments of racial stereotypes currently prevalent in the United States. Sowell (1994) addressed race-based stereotypes worldwide. See, e.g., Faludi (1991), Fausto-Sterling (1992), and Wolf (1991) for popular treatments of gender stereotypes in the United States.

2.1 INFLUENCES FROM "SOCIAL PSYCHOLOGY OF LANGUAGE"

The field of social psychology of language has laid some of the ground work for the perception research that I discuss here, and so it is worth briefly mentioning at this point. Researchers in the field of social psychology of language (which draws heavily from social psychological theory and is closely related to sociolinguistics and sociology of language[2]) have for some time been strong proponents of the idea that the way in which individuals "assess" others linguistically affects further judgments made about those individuals. The notion that stereotypes affect the general perception of people is well-documented and wholly accepted within the field. Listeners react to speakers based on socially-influenced beliefs about the language that the speakers use. That is, linguistic stereotypes affect the way that people are perceived in a social context.

Nearly four decades ago, for example, social psychologists who were interested in testing the effects of such linguistic stereotypes on listeners' assessment of speakers' social characteristics developed the "matched guise" technique (Lambert, Hodgson, Gardner, & Fillenbaum 1960; Lambert 1967). In this technique, listeners from the same speech community hear speech samples from a number of speakers and then judge the speakers on a variety of personality characteristics which are of interest to the researchers. Wardhaugh (1992:113) listed common examples of qualities examined in matched guise studies as being: "intelligence, kindness, dependability, ambition, leadership, sincerity, and sense of humor." These qualities are assessed using semantic differential scales (e.g., a 7-point scale for intelligence, with "7" being "very intelligent" and "1" being "not intelligent,"

---

2 Chambers (1995:11) differentiated the foci of these two fields by stating that sociology of language is "the study of the relationships between language and society with the goal of understanding the structure of society," while sociolinguistics is "the study of the relationships between language and society with the goal of understanding the structure of language." Or, to restate it in Hudson's (1980:4-5) earlier words, sociology of language is "the study of society in relation to language," whereas sociolinguistics is "the study of language in relation to society." Same evidence, slightly differing objectives.

where the listener marks the point on the scale deemed appropriate for the speaker). The degree to which the speaker is perceived to personify each quality is then calculated (see Fasold 1987:150-151 for a brief explanation of this calculation).

The illuminating feature of the matched guise technique, however, is that the speakers employed in the experiment are multilingual, or multidialectal as the case may be. Listeners hear the speaker not once, but several times, each time using a different language variety (or "guise"). By having a single speaker alternate between several varieties (the content of the speech itself remains the same in both trials), the listeners are induced to react to the language *variety*—while controlling for individual speaker voice characteristics and speech content—when they use the semantic differential scales to assess the personality characteristics of the speaker. Results from the scales are compared for each guise of a single speaker's performance to assess each variety's influence on perception of personality characteristics.

However, it is important to see how language attitude studies of this nature quantify the folk linguistic beliefs that all speakers hold, both about their own and other varieties of speech. Such studies "clearly indicate that listeners are affected by code choices when they judge what speakers say to them" (Wardhaugh 1992:114). That is, social stereotypes about language clearly affect the way that speakers are socially perceived (see Preston 1999 for a review of current related work). But what this field of research has not considered is that these same social stereotypes actually play a role in shaping *the perception of the speech itself*, which is an important point for consideration in the present work.

Thakerar and Giles (1981) offered an early review of studies that focused on the effects of speech cues on the formation of various social judgments. They cited, for example, evidence that Standard American English speakers tend to be upgraded on competence and status-related traits, that slower talkers are rated as less competent, and that African American children are consistently evaluated by teachers as possessing poorer verbal abilities in general than Anglo students are viewed to possess. Thakerar and Giles generalized that social beliefs evoke speech stereotypes of how individuals are expected to talk.

Recent work reported in Purnell, Idsardi, and Baugh (1999) showed how social judgments about race/ethnicity formed on the basis of speech can lead to unlawful housing discrimination. This work was based on the difficulties that Baugh himself experienced when he tried to rent an apartment in the San Francisco Bay Area several years ago (Baugh is African American). He wondered whether the landlords' determination of his race/ethnicity, via his voice and dialect, was a factor in his being denied numerous appointments to visit apartments for rent. When he contacted the landlords on the telephone to request appointments, in many cases he would be told that the advertised apartment had already been rented, when in fact he knew that it had not.

As Baugh is tri-dialectal in African American Vernacular English, Chicano English, and Standard American English (his terms for each of the varieties, or "guises"), he investigated this issue by making three telephone calls to each of a number of prospective landlords to ask for appointments to see apartments, with all calls to a single landlord made 30 minutes apart and once in each guise. The pattern of the landlords' responses to Baugh in his various guises indicated that discrimination in opportunities for housing does indeed occur, as triggered by perceived dialect and ethnicity: landlords on average across the Bay Area were least likely to grant Baugh an appointment when he used Chicano English, and most likely when he used Standard American English. (See Baugh 1999 for an on-line presentation of samples of the speech stimuli in each of his three guises.)

Daly and colleagues (Daly, Bench, & Chappell 1996; Daly 1997) examined the impact of gender stereotypes on speechreading.[3] Daly et al. (1996) found that interpersonal impressions are also formed in non-auditory, visually-read speech, in ways similar to those of auditory speech. Their perception experiment results indicated that talkers who are judged to be more intelligible are also rated more positively interpersonally and that male talkers are generally rated higher on status measures. They also found that female talkers are generally rated higher on performance, which the authors hypothesized was the result of the commonly-assumed stereotype that women hyperarticulate their speech. The authors concluded that the perception of visual speech is affected both by the speech variables used by talkers as well as the gender stereotypes held by speechreaders. They further proposed that what they term "broad-level societal influences" (in this case, gender stereotypes) influence not only the speech variables used by talkers, but also the speech perception of speechreaders.

In their review on the history of research on Speech/Communication Accommodation Theory, Giles, Coupland, and Coupland (1991) discussed a number of studies of speech accommodation which suggest that stereotypes do play a role in communication, in that speakers adjust their speaking styles in relation to social perceptions about their interlocutors. The authors summarized that "accommodation strategies characterize realignments of code," with speakers adjusting their own production of speech in interactions based on beliefs and expectations that they have about the people with whom they are interacting. They cited Beebe (1981), for example, who studied the language of Chinese Thai bilingual children who were interacting with adult speakers of Standard Thai. Beebe found that with adult speakers who looked ethnically Chinese, the children used

[3] Daly et al. (1996:27) defined *speechreading*, commonly known as lipreading, as "a form of communication used primarily when the auditory signal is reduced or degraded in some way, usually by hearing loss or competing noise, and hence when hearing alone does not permit adequate perception of the speech signal."

Chinese phonological features. As Giles and colleagues stated, in language usage, "speech accommodation is often cognitively mediated by our stereotypes of how socially categorized others will speak" (p. 16).

Clearly, stereotypes based on language use influence social perception, a notion that has been a cornerstone of work in the area of social psychology of language study, since the field's inception. The work that I describe here draws from this notion, but emphasizes instead the way in which these evaluative judgments actually play a role in shaping the perception of language itself.

2.2 CONCEPTUALIZING STEREOTYPES

Before embarking on a discussion of what stereotypes are and how they are theorized to fit into cognition, I will first offer a brief description of how they have been conceptualized throughout the short history of social psychology. As I will discuss later in this chapter, the related notion of the *perceiver as cognitive miser*, which developed from social psychological work in the 1970s, is particularly salient to the argument that perceivers incorporate social stereotypes into the speech recognition process.

The scientific study of stereotypes and stereotype processes has traditionally fallen within the realm of social psychology, a field that applies scientific method to the study of the way individuals think, feel, desire, and act in social situations (Brehm & Kassin 1996:6). Since the development of the concept of information processing a half-century ago (which refers to the notion that mental processes can be broken down into consecutive stages, and thereby counters the earlier behaviorist notion that the only valid objects of study are the observable stimulus and response), "social psychologists have decided that social behavior is more usefully understood as a function of people's perceptions of their world, rather than as a function of objective descriptions of their stimulus environment"

(Fiske & Taylor 1991:9). Thus, in an information processing-view of perception, there is room for stereotypic beliefs not reflected in the objective stimulus environment to affect the thoughts, feelings, and actions of people.

Fiske and Taylor (1991:9) offered the following example to further illuminate this notion:

> . . . an objective reward like money or praise that people perceive as a bribe or as flattery will influence them differently than a reward they perceive as without manipulative intent. What predicts their reaction, then, is their perception, not simply the giver's action. Other people can influence a person's actions without even being present, which is the ultimate reliance on perceptions to the exclusion of objective stimuli. Thus, someone may react to a proffered bribe or to flattery by imagining the reactions of others ("What would my mother say?" "What will my friends think?"). Of course, such thoughts are the person's own fantasies, having perhaps tenuous connection to objective reality. Thus, the *causes* of social behavior are doubly cognitive; our perceptions of others actually present and our imagination of their presence both predict behavior.
>
> [original authors' emphasis]

Such research on social behavior was begun nearly a century ago when the field of psychology, just then developing out of the longer-historied area of philosophy, came to intersect with sociology (see Jones 1998 for a full review of the history of the field's development). The general focus of research at this intersection was the behavior of the individual (the target of psychological study) in the context of the surrounding group (the target of sociological study). Early research in social psychology aimed simply to measure and describe the content of various stereotypes and prejudices.

The field of social psychology and stereotype research has always been very much influenced by salient social events and issues of the time. In the aftermath of World War II, for example, social psychologists refocused their research in an effort to explain the emergence and power of Hitler's Nazi regime and the anti-Semitism surrounding the Holocaust. The result was the development of the Authoritarian Personality Theory

(Adorno, Frenkel-Brunswik, Levinson, & Sanford 1950), which addressed the "pattern of hatred of out-groups and a particular character structure responsible for these entrenched prejudices" (Fiske 1998:358).

With the 1960s came increased awareness and visibility of social injustice in the United States. Following this, social psychologists again refocused their research efforts, this time to a new interest in covertly-expressed racism and sexism. In terms of racism, the resulting Subtle Racism Theory addressed issues such as the fact that even though *reported* racial attitudes had changed dramatically in the United States over the past several decades, actual *behavior* had not changed much (e.g., given the opportunity to help a white stranger or a black stranger, whites were more likely to help whites, as well as give better help than they gave to blacks [Fiske 1998:359]). Various measures of ambivalent racism as well as sexism are the result of this era's research. (See Fiske 1998 for a review; also see McConahay & Hough 1976 for a description of their Modern Racism Scale; Swim, Aikin, Hall, & Hunter 1995 for a description of their Modern Sexism Scale; and Tougas, Brown, Beaton, & Joly 1995 for a description of their Neo-Sexism Scale.)

During the late 1970s and into the 1980s, social psychologists began to more directly examine stereotyping in its relation to other cognitive processes in an information processing view, and came to consider stereotyping as merely a by-product of normal cognitive processes, especially categorization. While earlier stereotype work on, e.g., authoritarian personality, had emphasized the abnormality of prejudice, these newer cognitive approaches saw stereotyping as just another component of the information processing system, and sought to explain it in that light (Fiske 1998:362). Stereotyping was now viewed as operating automatically (i.e., prior to conscious processing) within the information processing system, and as a component serving to aid in the categorization of newly-encountered stimuli. (Deaux 1995:13 cited Devine 1989; Dovidio, Evans, & Tyler 1986; and Gaertner & McLaughlin 1983 as addressing stereotypes' automatic operation.)

Notions of categorization and the way that the human cognitive system handles categorization were crucial to the consideration of stereotype effects during this period of research. As researchers had recognized for some time, people naturally categorize other people, as well as objects, during their daily course of being bombarded by incoming stimuli. Fiske (1998:361) quoted Allport (1954:175-176) as follows:

> Impressions that are similar, . . . especially if a label is attached, . . . tend to cohere into categories (generalizations, concepts). All categories engender meaning upon the world. Like paths in a forest, they give order to our life-space . . . The principle of least effort inclines us to hold to coarse and early-formed generalizations as long as they can possibly be made to serve our purposes . . . An ethnic prejudice is a category concerning a group of people, not based on defining attributes primarily, but including various "noisy" [nonessential, possibly false] attributes, leading to disparagement of the group as a whole. [information in square brackets added by Fiske]

Building on Allport's earlier notions of categorization in social contexts, stereotyping was now viewed as an entirely natural part of every human's cognitive system of processing information. It is at this point in social psychological history that the notion of the *perceiver as cognitive miser* (Taylor 1981) was developed, which holds that people, as perceivers of a never-ending incoming stream of stimuli, are "overwhelmed by the complexity of the social environment and [are] forced to conserve scarce mental resources" (Fiske 1998:362). This conservation is done presumably by employing the cognitive shortcuts that pre-established perceptual categories and their related attributes afford the perceiver. So, not only were stereotypes now viewed as natural aspects of the cognitive system, but they were also recognized as being very useful.[4]

Fiske and Taylor (1991:13) suggested that the idea behind the cognitive miser model is that:

> . . . people are limited in their capacity to process information, so they take shortcuts whenever they can . . . People adopt strategies that simplify complex problems; the strategies may not be normatively correct or produce normatively correct answers, but they emphasize efficiency. The capacity-limited thinker searches for rapid adequate solutions, rather than slow accurate solutions. Consequently, errors and biases stem from inherent features of the cognitive system, not necessarily from motivations.

The categorization that serves as such a cognitive shortcut is theorized by Taylor (1981) to work in several ways (as summarized by Fiske 1998:362-363):

> . . . categorization (1) tags information by physical and social distinctions such as race and gender, (2) minimizes within-group differences and exaggerates between-group differences, and (3) causes group members' behavior to be interpreted stereotypically. As a result of categorizing a set of people into two or more groups, smaller groups (i.e., solos, pairs, or minorities within a larger group in any given setting) elicit (4) fewer distinctions among themselves and (5) more stereotyped perceptions. Increasing familiarity, however, (6) allows more distinctions and (7) creates subtypes.

---

4 Huxley's (1954) essay "The Doors of Perception" is an interesting early example of the notion of "cognitive miserliness"—in this essay, Huxley gives a detailed account of the effects that he experienced one evening after ingesting the hallucinogenic drug mescaline, which is the active principle in the desert cactus known as *peyote*. Huxley found that the drug, which is chemically very similar to adrenaline, greatly increased his sensitivity to aspects of stimuli of which he would otherwise have remained unaware. This experience prompted him to support an early theory of cognitive miserliness, in that "the function of the brain and nervous system and sense organs is in the main *eliminative* and not productive. Each person is at each moment capable of remembering all that has ever happened to him and of perceiving everything that is happening everywhere in the universe. The function of the brain and nervous system is to protect us from being overwhelmed and confused by this mass of largely useless and irrelevant knowledge [sic], by shutting out most of what we should otherwise perceive or remember at any moment, and leaving only that very small and special selection which is likely to be practically useful" (pp. 22-23, original author's emphasis).

More recent developments in the understanding of social cognition again focus on the importance of motivation as well as emotion in information processing, which the earlier construction of the perceiver as simply a cognitive miser could not always account for. Adapting the notion of the cognitive miser to account for a more complex set of observed experimental occurrences has led to the more accommodating view of the social perceiver as a *motivated tactician*, which Fiske and Taylor (1991:13) described as:

> . . . a fully engaged thinker who has multiple cognitive strategies available and chooses among them based on goals, motives, and needs. Sometimes a motivated tactician chooses wisely, in the interests of adaptability and accuracy, and sometimes the motivated tactician chooses defensively, in the interests of speed . . . Thus, views of the social thinker are coming full cycle, back to appreciating the importance of motivation, but with increased sophistication about cognitive structure and process.

In short, modern conceptions of social perception and stereotyping focus on effects within a social context, and assume "fundamental cognitive processes of categorization and other cognitive economics" (Fiske 1998:364).

2.3 DEFINING STEREOTYPES

As Fiske (1998) reviewed, there are perhaps as many definitions for the notion of *stereotype* as there are stereotype researchers, but a basic chord that runs through all definitions is that stereotypes are cognitive structures that consist of "beliefs that people have about individuals based on their membership in social groups" (Brehm & Kassin 1996:122). Hilton and von Hippel (1996:240) further elaborated the notion of stereotypes as beliefs about the attributes of groups by adding that stereotypes are also "theories about how and why certain attributes go together."

The tendency for generalization in stereotyping is important to emphasize, as well; as Snyder and Miene (1994:34) stated, "stereotypes are usually simple, overgeneralized assertions about what 'they' are like, 'they' being the members of social categories who are robbed of their individuality by having applied to them a set of beliefs that ascribe to them, one and all, a set of shared attributes of character and propensities of behavior." It is important to mention at this point that these "overgeneralized assertions" can be based on direct observation (the interpretation of which, of course, might not necessarily objectively represent the stimulus), as well as indirect observation or hearsay.

Additionally, crucial to an understanding of the cognitive representation and function of stereotypes are the points that Brewer (1994:317) summarized:

> . . . [within the tradition of social cognition research,] stereotypes are conceived of as category-based expectancies that influence attention, interpretation, and memory for information about social category members (Hamilton et al. 1990). Category stereotypes originate in the service of cognitive economy, and biases in information processing, storage, and retrieval are understood as consequences of normal cognitive functioning.

The origins of social stereotypes come from a number of sources (see, e.g., Allport 1954). Brehm and Kassin (1994:122), for example, identified three such sources as being historical, political, and sociocultural in nature, and gave the following examples of each. Historically, stereotypes can arise from past events. "It can be argued that slavery in America gave rise to the portrayal of Blacks as inferior, just as the sneak attack on Pearl Harbor in World War II fostered a belief that the Japanese cannot be trusted." Politically, stereotypes can be a "means by which groups in power come to rationalize war, religious intolerance, and economic depression." And socioculturally, "it has been argued that real differences between social groups contribute to the birth of perceived differences."

Regardless of the origin, several theories have been put forth to explain the formation of stereotypes (see, e.g., Hilton & von Hippel 1996 for a general review of stereotype theories as well as the cognitive and motivational factors that contribute to stereotype formation, maintenance, application, and change). Brehm and Kassin

(1994:122-123) hypothesized that stereotype formation involves two related processes, the first of which is the categorization of others on the basis of common attributes that are socially salient within a given culture.

In American culture, for example, as well as many others, these attributes include sex/gender, age, and race. (In many social psychological studies, sex/gender is often shown to be the most conspicuous, immediate social categorization made by perceivers—Brehm & Kassin 1994:139 cited Stangor, Lynch, Duan, & Glas 1992 and Zarate & Smith 1990 as examples of such studies.) One serious drawback of such social categorization, however, is that we as perceivers tend to overestimate the differences between groups while at the same time underestimating the differences within groups.

After socially categorizing others according to common attributes comes their identification as either ingroup or outgroup members. The notion of the *outgroup homogeneity effect* follows directly from this group membership identification, and refers to the phenomonenon that outgroup members are perceived as more homogeneous to one another than are ingroup members. Outgroup members are also perceived to possess generally less-desirable traits than are ingroup members.

Additionally, minority groups are perceived as more homogeneous than majority groups (Hilton & von Hippel 1996:247). The outgroup homogeneity effect is thought to occur both because perceivers often have relatively little contact with outgroup members (thus making it more difficult to perceive actual differences first-hand), and that perceivers also are not likely to encounter a representative sample of outgroup members in the contact they *do* have with them. As Brehm and Kassin (1994:124) pointed out, the outgroup homogeneity effect is common:

> Americans who arrive from China, Korea, Taiwan, and Vietnam see themselves as different, but to the western [non-Asian] eye they are all Asian. Business managers like to talk about engineering types, engineers

> talk about business types, liberals lump together all conservatives, teenagers lump together all old people, and as the natives of New York City proclaim their cultural and ethnic diversity, outsiders talk of the typical New Yorker.

Self-fulfilling prophecies are another often-considered explanation for stereotype formation. This notion is straight forward: "self-fulfilling prophecies emerge when people hold expectancies that lead them to alter their behavior, which in turn causes the expected behaviors to be exhibited by people who are targets of the expectancies" (Hilton & von Hippel 1996:244). As Fiske and Taylor (1991:545) explained, several critical steps must take place for a self-fulfilling prophecy effect to occur: "The perceiver must hold an expectation about a target, and behave in a manner consistent with it. The target then interprets the behavior, and responds to the perceiver with actions that are subsequently interpreted by the perceiver as consistent with the original expectation."

Illusory correlation is an additional phenomenon that has been examined as a basis for stereotype formation. Illusory correlation refers to the tendancy for people to overestimate the link between variables that are only slightly or not at all correlated (Brehm & Kassin 1994:125). Fiske and Taylor (1991:373) cited the classic illusory correlation study, a semantic effect described in Chapman (1967):

> . . . college-student stubjects [were presented with] lists of paired words such as lion-tiger, lion-eggs, and bacon-eggs. Subjects were then asked to report how often each word was paired with each other word; in fact, all words were paired with each other an equal number of times. Nonetheless, subjects reported that words that were associated in meaning such as bacon-eggs had been paired more frequently than those not associated in meaning.

Carrying this notion into the realm of social perception, Hilton and von Hippel (1996:245) pointed out that illusory correlation highlights people's tendency to perceive minority groups more negatively than they do majority groups, even when the groups behave in the same way. One explanation for this is that minority groups are perhaps viewed as more distinctive in the first place, as are negative behaviors, and the two come to be associated with each other at the time of encoding through their shared distinctiveness. But no matter how illusory correlation actually works, it is important to point out that stereotypes formed through illusory correlations are generally quite robust and resistant to change.

2.4 Representing Stereotypes

In addition to considering the cognitive processes that lead to stereotype formation, social psychologists are also interested in modeling the ways that stereotypes might be mentally represented, maintained, applied, and changed. And as I will be making the point in later chapters that stereotypes are operative at very low levels of cognition, it is necessary here to talk about stereotype representation.

There are a number of ways that the representation of stereotypes in memory has been modeled, including stereotypes as associative networks, stereotypes as schemas, stereotypes as prototypes, and stereotypes within exemplar frameworks. Stereotypes have been most commonly represented within either associative network models or schema models, and prototype models have often been used to represent their architecture. But some social psychologists are beginning to consider stereotype representation within exemplar models, as well as how they might be treated within a parallel distributed processing (PDP) connectionist architecture.

Smith (1998:391) defined a *representation* in psychological terms as "an encoding of some information, which an individual can construct, retain in memory, access, and use in various ways." And Fiske and Taylor (1991:296-297) defined *encoding* as "the process of taking in data and creating a mental respresentation of it." In order to give some helpful structure to the discussion of these different, yet related, types of representational models, Smith (1998) separated them into two general metaphorical categories: associative networks, schemas, and prototypes were viewed as representations of "things"; while exemplars and PDP mechanisms were viewed as representations of "states" and are intrinsically time-bound in Smith's conceptualization.[5]

Smith suggested that his metaphors imply different general properties for each type of representation. Representations as "things," for example, are searched for in memory, with the representations that are not selected as the result of the search then having little contribution to the current process. On the other hand, a supporter of an exemplar or PDP model would contend that "all representations are part of the person's overall state, so no representations remain unused or inert; any representation might influence current processing" (Smith 1998:392).

Associative network models assume networks of discrete nodes, which are connected (or "associated" to each other) by different types of links. Hilton and von Hippel (1996:242) pointed out that in stereotype representation, the networks are linked attributes which have been variously defined as either traits, beliefs, or behaviors of social groups (Fiske & Taylor 1991 and Smith 1998 also offered detailed descriptions of associative network models). Such models predict that these related ideas, traits, beliefs, or behaviors (the "nodes") will be recalled together (via their "links"); when a node is sufficiently activated, the attribute will be retrieved from memory. Links between attributes are strengthened when the attributes are either experienced or thought about in conjunction with each other, i.e., each time the link is activated by consideration of the nodes it connects.

Another important aspect of associative networks is that links are "labeled," meaning that the links maintain knowledge about the proper relationship between nodes. These models hold that associations vary in strength and can be activated automatically and

[5] Smith's sub-classification of these models of mental representation is a befitting example of

without the perceiver's conscious consideration (i.e., implictly), and that stereotypes modeled as strongly interlinked groups of nodes in associative networks change very slowly due to the high degree of interconnectivity between attributes that make up the stereotype (Hilton & von Hippel 1996:243).

In summary, associative networks are considered to be composed of associative links between nodes bound together into a single representation of the different perceived aspects of an object or concept, often including related information such as the specific context in which the object was encountered. The binding occurs when aspects are thought about together, and gives a "unified representation that can be retrieved as a whole when activated by a part" (Smith 1998:402). Associative networks are assumed to operate in a "bottom-up" fashion, or are strictly stimulus- or data-driven. One of the criticisms of this model is that many experimental findings suggest that perceivers actually use multiple processing strategies, depending on the circumstances, which cannot be captured by the pure associative network structure (see Fiske & Taylor 1991:306).

Schematic representations, in contrast to associative networks, are higher-order, abstract representations with significant internal structure (Smith 1998:403). Fiske and Taylor (1991:98) defined schemas as "cognitive structures that represent knowledge about a concept or type of stimulus, including its attributes and the relations among those attributes . . . schemas are concerned with the general case, abstract generic knowledge that holds across many particular instances." Likewise, schemas are thought to facilitate "top-down, conceptually-driven, or theory-driven processes, which simply means processes heavily influenced by one's organized prior knowledge, as opposed to processes that are more bottom-up or data-driven," which associative networks are considered to facilitate.

---

categorization as an aid to perception!

Schema theories try to explain how perceivers "understand and remember complex material by drawing on abstracted general knowledge about how the world works" (Fiske & Taylor 1991:103). Following this, schemas are usually considered to act at a preconscious level (Smith 1998:403). Within schema models, stereotypes are often considered to be essentially "role schemas," or sets of expected behaviors or attributes of persons in particular social positions or categories (i.e., "roles"; Fiske & Taylor 1991:119).

Unlike the tightly linked groups of nodes in associative network structures, schemas are considered to be independent entities; when one becomes activated, there are no necessary implications for other related schemas. As Hilton and von Hippel (1996:243) stated, in a schema view of stereotypes:

> [the] stereotype of men, for example, may contain the general belief that they are agressive, without tying that belief to particular instantiations of agression, specific contexts or exemplars, or particular organizational structures.

Schemas are activated by explicit thought about the category, or by an encounter with real-world relevant stimuli, and the activation of a schema makes all structured knowledge contained within the schema accessible. The primary function of activated schemas is interpretive: schemas influence the evaluation of incoming stimuli. Additionally, schemas can serve as a "fill-in guide" for perceiving otherwise-ambiguous stimuli. Following from the Gestalt psychologists who believed that perceptions are mediated by context, schema representations address the notion that "our perceptions of the world reflect an interplay between what's out there and what we bring to it . . . As structured knowledge that we bring to everyday perceptions, schemas emphasize our active construction of reality . . . We experience the world as if our schemas have added nothing to it, so common sense tells us that we perceive an unchanged or literal copy of the environment" (Fiske & Taylor 1991:99).

One important principle in schema operation is that before a perceiver can apply a schema to the interpretation of a target, that perceiver must first classify the target according to categories held by the perceiver. Fiske and Taylor (1991:105) offered this example:

> Once you recognize someone as filling a particular role (e.g., gas station attendant) on the basis of particular attributes (probably male, maybe wearing a jumpsuit, maybe with an oily rag in pocket, maybe approaching your car, perhaps helping others get gas), then you can apply your knowledge about the role to guide the subsequent interaction.

So how is it that perceivers perform such categorization? The process of categorization in general is of great interest (see Fiske & Taylor 1991:105-107 for a short review of categorization theories, and Rosch 1978 for a classic treatment), but one way that social categorization (and stereotyping) has been theorized to occur is addressed by prototype models.

Prototype models usually assume an abstract representation that is averaged across the typical attribute values of the members of a category, with similarity comparisons of individuals made to these averages. "Social categories thus may be viewed as fuzzy sets centering on a prototype" (Fiske & Taylor 1991:107). Prototype models employ a hierarchical organization structure in which subtypes can be formed; the categories in these hierarchical organizations become more inclusive as they become broader.

Prototype models are also considered to include all known attributes of the category, even though none are actually defining of the cateogory, while schema models are considered to contain only category-relevant attributes. Stereotypes in a prototype model, then, are simply abstracted representations of a specific group's typical features, with individuals judged according to how closely they match the prototype (Hilton & von Hippel 1996:241).

Hilton and von Hippel (1996:242) noted that one prediction of this type of model is that people may often fail to apply a certain stereotype to an individual if too much disconfirming information is present:

> Because reactions to individual group members are based on a comparison between the prototype and the individual, any features . . . that reduce the similarity between the individual and the prototype should decrease reliance on the stereotype.

One criticism of prototype models is that categories might not always be best represented by an averaged prototype, but might more often be represented by ideals or extremes (Fiske & Taylor 1991:109). A further critique of the prototype model is that the averaged representation doesn't actually represent the category as well as the entire collection of experienced examples would. There are a number of experimental results that show, for example, that perceivers retain and use detailed information about group variability; simple prototype models cannot account for this knowledge about variation.

Exemplar models, on the other hand, store specific examples to represent groups. These examples are episodic memories of the actual stimuli that the perceiver perceived, as opposed to generic memories of "features" or other information abstracted away from the stimulus. The function of exemplars is essentially the same as that of schemas: to influence perception of an incoming stimulus. Exemplar models emphasize context and goals involved in a specific encounter, which allows different stereotypes to be activated and applied to various group members on different occasions.

In terms of activation, all exemplars are activated in parallel by retrieval cues, which can be either generated by thought or through an actual encounter with an external stimulus. The level to which each individual exemplar is actually activated depends on its similarity to the retrieval cue (the more similar, the higher the level of activation). Representations of stimuli in exemplar models can be thought of as "states," to return to Smith's metaphor, because such representations are "temporary, on-line [constructions] formed from concurrently activated exemplars, which will change over time and in different contexts" (Smith 1998:418).

Smith (1998:412) summarized Smith and Zarate's (1992) exemplar-based model of social judgment and the way it handles the modeling of stereotypes (which seems a bit more abstract than the way that stereotypes can be represented in, e.g., a schema approach):

> . . . this model assumes (a) as the perceiver encounters or considers individual persons, cognitive representations of those persons (i.e., exemplars) are constructed and stored in memory. A representation of a "typical member of Group X" constructed on the basis of reflection about the group's general characteristics, a secondhand description of the group, or a joke or slur that implies the group possesses some stereotypical attribute, could also be stored as an exemplar. (b) When a judgment concerning a person or group is required, the known attributes of exemplars that are retrieved from memory by virtue of their similarity to the target form the basis of judgment. (c) Similarity is modulated by the perceiver's attention to stimulus dimensions, as in Medin and Schaffer's (1978) context model. For example, if in evaluating a female lawyer the perceiver pays attention to gender, then other representations of females would be treated as most similar and would be most likely to be retrieved—resulting, presumably, in judgments that the target possesses gender-stereotypic attributes. In contrast, attention to occupation would result in the preferential retrieval of other lawyer exemplars and, presumably, judgments that the target shares their attributes. And if the perceiver pays attention to the person's unique individual characteristics (e.g., her manner of speaking), still other representations (e.g., representations of others who speak in a similar way) will be retrieved to influence judgments about her. (d) Social and motivational factors, including the current social situation, the perceiver's self-concept,[6] and in-group/out-group dynamics, affect attention to stimulus dimensions and hence exemplar retrieval and judgment processes.

---

[6] *Self-concept* (also called "self-schema" in the literature) refers to the way that we categorize ourselves, in terms of the social categories that are salient to us (this notion was first introduced in Markus 1977). The effect of self-concept on perception, memory, and inference is tied to the notion that we categorize ourselves as much as we categorize others, which then influences information processing about others. "For example, people who consider themselves religious respond in more detail to religious (rather than, say legalistic) arguments on any given topic" (Fiske & Taylor 1991:184; see Fiske & Taylor 1991:181-195 for a review of the effects of the representation of self, including the self-concept).

Presumably, then, a stereotype effect within an exemplar model would be the influence of changes in attention weights during activation of the relevant exemplars prior to the comparison phase.

Parallel distributed processing (PDP) mechanisms are relatively new approaches to modeling cognition, and have been the least-used of the models that I have mentioned thus far in representing stereotypes, but many researchers see them as a promising avenue for exploration and for that reason I mention them here. In a PDP architecture, a concept or object is represented by what is termed a *distributed representation*, which is a pattern of activation across a large set of processing units in a module.

Unlike in an exemplar model, in a PDP model the processing units are not memories of individual examples or of anything else that is interpretable on its own. Rather, they are switches that can be turned on or off and that can then turn on or off other coupled switches to represent larger dynamic patterns. Smith (1998:419) gave the example of pixels in a television screen: each pixel on its own is essentially meaningless, but when activated in a certain way in conjunction with the other pixels on the screen, the result is a unified, meaningful representation.

All memory units in a PDP model are connected and activate each other across weighted connections which are uni-directional (versus multi-directional links in, e.g., associative models, which can spread activation both ways). Activations can either be positive or negative, thus having either a facilitative or inhibitory effect on a connected unit.

While associative networks, schematic models, prototype models, and exemplar models assume static representations and require additional assumptions about processing, PDP models process and store information in a single mechanism, with the entire pattern of activiation representing meaning (Smith 1998:419). "In a PDP model, each unit helps to represent many different concepts, which are [reconstructed] when the appropriate pattern of activation occurs across all the basic units" (Fiske & Taylor 1991:309). (But of course, a stereotype in a PDP model might be a partially "crystalized" pattern—i.e., a "local minimum" kind of allocation of specific subparts of the network—as when an image that stays on the television screen too long may "burn" a trace in the cathode-ray tube [Mary Beckman, personal communication].)

Representations are considered to be reconstructed rather than retrieved (as is the case in the other aforementioned models), and "only the strengths of connections are stored, so that the pattern can be recreated by activating parts of it and waiting for the connections to reverberate throughout the system until the entire pattern is activated" (Fiske & Taylor 1991:310). While other models may assume a distinction between top-down (knowledge-driven) and bottom-up (stimulus-driven) processing, PDP models assume that the two types of processing are actually inseparable since all processing here is performed by flows of activation (ultimately arising from perceptual input) through connections (which encode existing knowledge; Smith 1998:427).

PDP mechanisms have not yet had much effect on social cognition research, but since they are able to handle multiple sources of information simultaneously (hence the term "parallel" in the name of the model), their possible application to stereotype-related work is recognized by many researchers. As Fiske and Taylor (1991:310-311) commented:

> . . . [one] possible domain of application in social cognition is to schemas, and in particular to how schemas simultaneously interact with each other. For example, combining one's knowledge about traditional Amish farmers and progressive Montessori teachers, one can imagine someone who occupies both roles by considering their shared "back to basics" perspective and their shared emphasis on patience. Moreover, one can imagine the person's likely response to novel issues (e.g., computers in the classroom). PDP models allow for such emergent properties of previous knowledge. It remains to be seen whether and how they will influence social cognition research.

In Table 2.1 below, I present a brief typology of the various approaches to stereotype representation in cognition that I have described in this section. In summary, in an associative network model, the category has no representation of its own, but is imbedded in the connections of the model. In the schema and prototype models, the input primitives are the same features/attributes as for the associative network model, but now the category is instantiated as a node of its own. Stereotype categories are defined in terms of their connections to features. In an exemplar model, the input primitives activate specific exemplars to different levels depending on the primitives' similarity to the target. Finally, in a PDP model, the category is represented by a pattern of activation across a set of nodes.

APPROACH	PERCEPTUAL PRIMITIVES: WHAT ARE THE LOW-LEVEL "INPUT" ELEMENTS, AND HOW ARE THEY RELATED?	RELATIONSHIP TO CATEGORY REPRESENTATION	RELATIONSHIP BETWEEN TOP-DOWN AND BOTTOM-UP PROCESSING
Associative Network	Features / attributes of signal associated by correlation	Stereotype is a particular set of connected features	Emphasizes bottom-up: activation of a stereotype means activation of all connected features; hence related stereotypes are activated together bottom-up
Schema, Prototype	Features / attributes of signal activate abstract categories	Stereotype is schematized as abstract category separate from input	Emphasizes top-down: activation of a stereotype influences the processing of further bottom-up stimuli, but does not activate other stereotypes
Exemplar	Features / attributes of signal serve as retrieval cues to activate specific exemplars to different degrees	Stereotype is state of activation of certain exemplars linked to a common higher-order category	Top-down and bottom-up are inseparable; stereotype is an aggregation of stored exemplars linked by common connections to higher-order category
PDP	Features / attributes of signal activate pattern to reconstruct representation	Stereotype is pattern of connection strengths	Assumes top-down and bottom-up are inseparable since processing is performed by flows of activation

Table 2.1: Typological chart of the approaches to stereotype representation that have been described in this chapter. The various approaches are listed in the left-hand column. The second column describes the perceptual primitives assumed by each approach, or what gets mapped between the raw input signal and the categories. The third column gives a description of the relationship between the perceptual primitives and the stereotype category, and the column on the far-right summarizes the implications for the partitioning of bottom-up and top-down processing in each model.

So, the representation of stereotypes in cognition can be handled in a number of ways depending on the theoretical orientation of the researcher. While the majority of social psychological work to this point has assumed a representation of stereotypes as *schemas*, this representation is inadequate for a number of reasons, as I discuss further in the final sections of this chapter. It becomes clear that an exemplar model approach handles the effects that we have uncovered better than does a prototype model.

2.5 ACTIVATING STEREOTYPES

In this section, I review the ways in which the activation of stereotypes has been treated in the literature. No matter which representational view of stereotypes is maintained, it is important to emphasize several processes that have a crucial impact on how stereotypes come to be activated and thereby allowed to influence information processing. These phenomena include priming, assimilation effects, and contrast effects, which I briefly discuss below.

Priming is defined as "the effects of prior context on the interpretation of new information" (Fiske & Taylor 1991:257). Countless studies from numerous fields of cognition-related research have supported the notion that the way we process information is strongly influenced by previous experiences. Hilton and von Hippel (1996:248) summarized this notion by stating that "prior experience determines what we see and hear, how we interpret that information, and how we store it for later use."

In processing new information, people tend to rely on categories that they have used frequently or recently (Fiske & Taylor 1991:145). Stimuli tend to be assimilated into primed categories. Priming of all sorts of information, including information regarding personality traits, race, gender, and age, is said to lock perceivers into a stereotypic frame of reference, one which can remain operative for a long time.

Priming can occur in a number of ways and have a number of effects on information processing. (See, e.g., Best 1999:177-184 for a brief discussion of semantic versus lexical priming in lexical access, as applied to the general organization of internal knowledge.) Stereotype researchers who are interested in studying the effects that primed categories have on processing generally agree in singling out these three factors: "recently, frequently, and chronically-encountered categories are more accessible for use, and they profoundly influence the encoding of stimuli" (Fiske & Taylor 1991:266). According to several schema models of stereotypes, a schema that is activated frequently will maintain higher activation levels overall, and frequent activation over a long period of time is thought to prime a schema to a state of "chronic accessibility" (Fiske & Taylor 1991:264-265; Smith 1998:408). Chronically accessible schemas allow the automatic encoding of information (according to the schema) in an implicit fashion.

Recent work also suggests that the influence of an event's priming on evaluations is strongest when the events are stereotype-congruent. Hilton and von Hippel (1996:250) provided this example:

> Because the cultural stereotype holds that African-American males are hostile and unintelligent, for example, a stereotyping perceiver is primed by the mere presence of an African-American to interpret his behaviors as consistent with those traits. For this reason, if he were to do something friendly but stupid, it would likely be regarded as stupid, whereas his unfriendly but intelligent behavior would likely be regarded as unfriendly.

Fiske and Taylor (1991:258-259) described a number of experiments that highlight a priming effect on gender stereotyping:

> For example, men who had just viewed a pornographic film went on to respond more stereotypically to a woman they encountered in an apparently unrelated context: their behavior was judged to be more sexually motivated, and later they initially remembered mainly her physical features rather than the interview; but these results held only for gender-schematic men, for

whom gender role is likely to be especially accessible[7] (McKenzie-Mohr & Zanna, 1990). Relatedly, women primed with family terms remember more accurately and judge more confidently the goals of a wife/mother target person, compared to a career-woman target or compared to neutrally primed subjects (Trzebinski & Richards, 1986, cf. Trzebinski, 1985). And rock music videos that are gender-role stereotypic seem to prime stereotypic interpretations of men's and women's interactions (Hansen & Hansen, 1988b).

Stereotypes, then, are essentially arrangements of memory for which any salient label will robustly evoke the rest of the stereotype percept throughout the various levels of representation within memory. And as gender categorization is so automatic and spontaneous in the initial perception of others, gender stereotypes are continuously primed and therefore become chronically accessible to the information processing system (Cross & Markus 1993). In short, this means that most if not all processing occurs within the scope of the perceiver's gender stereotypes.

The notion of an *assimilation effect* refers to the phenomenon that individuals are often assimilated to primed categories, and therefore are perceived as more similar to the stereotype than they really are. Hilton and von Hippel (1996:250) noted that:

> . . . a student athlete is more likely to be judged guilty of cheating than a nonathlete, an angry housewife is seen as less aggressive than an angry construction worker, and an African-American pan-handler is seen as more threatening than a Caucasian pan-handler.

---

[7] The notion of *gender schematicity* is related to the concept of the self-concept (or self-schema) as I described in Footnote 6. Essentially, gender-schematic individuals are considered to be "masculine males and feminine females [who] are more attuned to these gender schemas and are more motivated to comply with them" (Fiske & Taylor 1991:170). So, people who are gender-schematic have a gender component in their self-concept which is more developed in the direction of norms for culturally-appropriate gender behavior, which then influences their information processing systems in ways that are different from how gender would affect processing if they were less gender-schematic (i.e., less stereotypical by the culture's standards) and thus less primed to perceive incoming information through such a prescribed gender lens. Mills (1983), for example, offered evidence for the ways that gender schematicity is theorized to influence processing. In this work, subjects first rated themselves on masculine, feminine, and neutral trait adjectives to determine their level of gender schematicity. In gender-schematic subjects, measures of self-ratings were more extreme, and when presented with sex role-congruent word pairs, response latencies were shorter and recall was better than for sex role-incongruent word pairs. Balanced individuals showed no difference.

Assimilation occurs to a greater extent when members of a stereotyped group are perceived as being more consistent with each other. Because outgroups are always perceived as more homogeneous than ingroups, outgroup members tend to be assimilated to the group stereotype to a greater extent.

The concept of a *contrast effect*, on the other hand, refers to the "tendancy to perceive stimuli that differ from expectations as being even more different than they really are" (Brehm & Kassin 1996:124). Brehm and Kassin pointed out that generally, judgments about incoming information are influenced by the amount of difference between the perceiver's expectations and the stimulus. If the perceived difference between stimulus and expectation is small, it is barely noticed. If the perceived difference between the stimulus and expectation is great, on the other hand, the difference tends to be magnified, and is thus a contrast effect.

Brehm and Kassin (1996:125) described research that examined how contrast effects can affect social perceptions:

> In a study by Melvin Manis and his colleagues (1988), subjects read sentences ostensibly written by mental patients at two hospitals and were led to believe that one group of patients was seriously disturbed but the other was not. Subjects then evaluated new sentences that depicted a *moderate* level of psychopathology . . . [results indicated that when] the material was supposedly written by disturbed patients, the [writers of the new] sentences seemed normal. When the same [original] sentences were thought to emanate from the more normal group, however, [the writers of the new sentences] seemed relatively disturbed. [original authors' emphasis]

Following this, as gender stereotypes tend to prime for "warm, gentle women and assertive, forceful men," individuals who violate these prescriptions beyond a certain degree will likely be subject to contrast effects in the perception of themselves by others (Brehm & Kassin 1996:125). Therefore, an outspoken woman may be perceived as a "bitch" and a gentle man may be perceived as a "wimp," when such behavior in actors of the other sex would elicit very different reactions.

The activation of stereotypes is thought to become automatic within the cognitive system over time, and likewise, the suppression of stereotype activation is difficult, "an effortful process that requires conscious cognitive resources from the perceiver" (Hilton & von Hippel 1996:254). Automatic stereotype activation is also robust in all perceivers, and has been shown to be cued by a number of information sources, including direct exposure to stimuli as well as indirect exposure through the discussion, suggestion, or thought of a particular person, object, or event. Activation of stereotypes seems to be accomplished with minimal cognitive resources.

2.6 Face-Induced Stereotype Effects

As I have mentioned above, stereotype researchers acknowledge that stereotypes can be activated through a number of "means of contact" with the stereotyped target, including both thought about the target and direct encounter with the target. For example, many researchers have examined the ways that seeing just the face of a person activates stereotypes about that person.

The face gender effect, as I will discuss in Chapter 3, essentially highlights the influence of the gender of a face of a talker on the perception of the talker's speech. In this case, seeing the face of a talker and categorizing talker gender (via the face) during information processing induces certain expectations about how the talker will speak, which affects the final speech percept. I wish to briefly consider the ways that appearance-based perceptions have been treated in the social perception literature, since some of the related theories may be applied to an explanation of the face gender effect.

Perhaps the most well-known body of research on the ways that faces induce stereotyped perceptions is the research on the *babyface effect*. (See Zebrowitz 1997 for a book-length treatment of all aspects of face-influenced social perception.) The basic result, first described by Berry and McArthur (1985), is that if someone is perceived as having babyish facial features, they are also perceived as being more childlike in their personality characteristics than are people with mature facial features; babyish-looking people are perceived as having babyish personal attributes (Taylor et al. 1994:282). Large, round eyes, high eyebrows, and a small chin, for example, yield perceptions of a babyish facial appearance, which is positively correlated with perceptions of naïveté, honesty, kindness, and warmth.

Zebrowitz-McArthur and Berry (1987), for example, have suggested that there is near-perfect cross-cultural agreement (in the cultures that they have studied to this point) about the traits that indicate babyfacedness, as well as resulting psychological traits, and that such stable perceptions of babyfaced adults derive from the "specieswide adaptive value of analogous reactions to real babies." Whether or not this is the actual cause behind the babyface effect, for the sake of the face gender effect it is valid to recognize that facial features have been shown to affect other types of processing (in this case, the attribution of otherwise unknown personality characteristics).

The way that sex/gender is perceived via facial information has also been examined. It has been shown, for instance, that babies learn to distinguish gender at very tender ages. Between the ages of 3.5 and 4 months, for example, infants become sensitive to audiovisual information specifying gender, and tend to look more at faces for which gender-appropriate voices are played (Kuhl & Meltzoff 1982, 1984; Walker-Andrews et al. 1991). Driver Leinbach and Fagot (1993) showed that by one year of age, infants have formed male and female categories in picture discrimination tasks of both faces and clothed, full body images.

And while people are able to identify the sex of a face with very high accuracy (96% for faces not including hair, as reported by Burton, Bruce, & Dench 1993; and 100% for the faces in our experiments, as I report in Chapter 4), several researchers have suggested that the facial features that are most salient for sex classification vary cross-culturally. For example, Yamaguchi and colleagues (1995) have suggested that for Japanese viewers, the cues that play the most important role in gender identification are the eyebrows and the face outline. On the other hand, Brown and Perrett (1993) have suggested that when Americans view composite faces, the most salient features for gender identification are the jaw, eyebrows and eyes, and chin, in that order.

At least some of this cross-cultural variation in cue salience can likely be attributed to differences in prescribed facial adornment, make-up practices, etc., as well as the increased gender salience of certain features which serve to "separate the men from the women" in the culture in question.[8] Since categorization of gender via face is such a robust, early-formed cognitive ability, it is not ill-conceived to posit that this process may further impact information processing.

Green and Ashmore (1998) proposed that more attention, in fact, should be paid to the influence of physical appearance in stereotyping, and cited three types of evidence from social cognition work that they believed support their suggestion. First, they cited Brewer (1988), who offered a person perception model that emphasized the importance of physical characteristics in perceiving others. "She conceptualizes category prototypes as 'picture-like' in format, and assumes that traits (ascribed attributes) are inferred from the pictoliteral representation (identifying attributes) of a category. According to Brewer, when an individual encounters another person, categorization involves matching the target's appearance with an internal analogue representation of the prototype's 'looks.'"

Second, they cited a number of studies which have supported the notion that physical appearance influences the identification of social categories (Fiske & Taylor

[8] One moment spent glancing at the male models in a *J. Crew* catalog, for example, yields evidence that our culture views a strong jaw line to be a marker of masculinity. It is not surprising, therefore, that the jaw has been shown by at least one group of researchers to be the most salient indicator of facial gender in American culture. Kiyoko Yoneyama (personal communication) corroborates this general notion for Japanese culture when she states that the most feminine representation of the face in Japan includes very

1991). Visual cues such as facial appearance and gender traits have been shown to influence stereotyping. And beyond physical appearance alone, several studies have also shown that imagery (i.e., beliefs induced by the suggestion of, e.g., occupation) plays a role in social stereotyping, which is supported in our own suggestion set experiment, as I describe in Chapter 3. Green and Ashmore cited Klatsky, Martin, & Kane (1982), who showed modifed faces to subjects for judgment on whether the faces were "natural" or composed of halves from different faces. Judgment reaction time was measured, and Kaltsky and colleagues found that it took a significantly longer amount of time to correctly identify a face as "natural" if it was preceded by a prime for an occupation which stereotypically did not match the face shown. (They gave the example of a "truck driver" prime followed by the face of a judge.) The researchers interpreted this finding "as indicating that the occupational label automatically elicted a mental image which then interfered with the processing of a visual stimulus that was incongruent with that image" (Green & Ashmore 1998:1611).

Finally, Green and Ashmore cited work such as that by Deaux and Lewis (1984), who found that physical appearance plays a dominant role in descriptions of gender categories. These three forms of evidence support their endeavor to qualify the mental images formed and used by perceivers, which is based on the notion that "people do not think of physical attributes as discrete bits of verbally encoded information (e.g., 'tall'), but rather that such knowledge is cognitively encoded in terms of an internal gestalt image (i.e., a mental picture; Brewer 1988) [which has functional importance in] encoding, comprehension, and recall" (Green & Ashmore 1998:1611-1612).

In their own work, Green and Ashmore assessed the mental images that people have of several culturally relevant gender types, considering three relevant factors: setting

---

finely-, almost extremely-plucked eyebrows; hence, perhaps, the identification of the eyebrows by of the gender type; clothing worn which indicates group membership in the type; and physical appearance (including body shape, face, and personal grooming). The gender types examined have repeatedly been identified as representing the way perceivers in American culture grossly categorize men and women. The four male types examined were the *business executive*, the *ladies' man*, the *homosexual*, and the *nerd* (as termed by the experimenters). The four female types examined were the *housewife*, the *whore*, the *career woman*, and the *feminist* (again, as termed by the experimenters; conceptual frameworks of each type are described on pp. 1614-1617).

Researchers asked participants to imagine and orally describe the way each type appeared in their mind, and the transcribed descriptions were then subjected to content analysis using 31 coding variables which had been developed empirically from a first-pass of the descriptions. Results were consistent across the college-aged participants involved, as well as across male and female subject groups, which further supports the consistency and saliency of these mental images. Analysis of the mental images for male types revealed that the business executive and ladies' man types were perceived to appear quite similar, as were the homosexual and nerd types, although the nerd type was viewed as slightly more unappealing. (In fact, Ashmore, Del Boca, & Wohlers 1986 reported that conceptions of the nerd and homosexual types were "located close to each other in a multidimensional space of male types, suggesting that perceivers conceived of them as similar social types" (Green & Ashmore 1998:1625).)

Analysis of the imagined female types indicated that the greatest similarities in appearance were between the housewife and the feminist, with the greatest divergence from the other three being apparent in the image of the whore. This research is the first to closely examine the mental images of culturally salient physical stereotypes, and thereby

---

Yamaguchi et al. 1995 as the most salient gender-by-face distinguisher in that culture.

attempt to identify the physical attributes that are stereotyped to accompany different gender types. While no techniques will ever likely make it possible to tap the "actual" images that a perceiver might hold in his or her head, this is a start on refining the methods necessary to better conceptualize this process.

2.7 Gender Stereotypes

Despite the nature of the "interaction" with an individual—i.e., whether that individual is discussed in absentia with an interlocutor, viewed in a photograph or videotape, heard on the radio or over the telephone, directly encountered in person, or merely thought about privately—the recognition of that person as an object for perception is enough for the cognitive system to activate any number of useful stereotypes which then serve to influence the information processing that commences in numerous ways. Additionally, in all cultures studied to this point, the identification of an individual's sex or gender is supported by a good deal of research as being one of the primary cognitive decisions made about an individual in any such type of interaction (see, e.g., Bem 1981). Deaux and colleagues (1985:145) stated it, somewhat simplistically, as follows:

> The human race can be divided rather easily into the two groups of males and females. A consequence of this fact is the development of cognitive categories to describe and process the gender-related information, a categorization process that can be termed "gender stereotyping."

It is worth briefly describing aspects of the general processes that constitute gender stereotyping, for these processes very likely have an effect on the perceptual results that I discuss in the chapters that follow.

While the definition of "gender stereotyping" as a "categorization process" is very helpful, I have characterized the preceding quote from Deaux et al. (1985) as "simplistic" because human perceivers do not use general stereotypes for just "man" versus "woman" in their processing of incoming information; people seem to be much more aware of and reactive to even subtle differences between individuals within gender groups. In a schema-based gender stereotype approach, the schemas that capture gender information and thereby influence perception should be fairly abstract characterizations of only the essential information necessary for category inclusion. This would follow from the earlier discussion of schema representation in Section 2.4; a schema is schematic.

This is difficult to reconcile with the fact that differences between individuals who have been identified as being members of the same sex categories have been shown to have a definite effect on further processing, unless the gender schema representations are more elaborate than just two schematic types. One elaboration is the notion of a "subtype."

In gender schema theory, the notion of *gender subtyping* has been a useful way to capture the sometimes-subtle differences between reactions to individuals, differences that have visible effects on the processing of other information about those individuals but would not be captured by a schema theory that recognizes only "female" and "male" as the most basic categories. Deaux and colleagues (1985:146) pointed out that several social scientists, following Rosch's (1978) model of categorization of natural objects, have suggested that the categories "male" and "female" may actually be superordinate categories, with the most salient gender-stereotypical categorization employing lower-level, more individual-specific gender categories (thus being, in Rosch's model, what are known as "basic" categories).

One of Rosch's basic suppositions is that the degree to which categories overlap in attributes serves as an index of the independence of the categories. "A minimal degree of overlap would support the contention of basic-level categorization" (Deaux et al. 1985:148). In this vein, Deaux et al. (1985:146) pointed out that:

> . . . just as objects in a living room are most meaningfully described by such terms as "chair" or "table" rather than the broad concept of "furniture,"

> perhaps people are seen mainly as "mothers" or "fathers" (or any number of other subcategories), rather than as "women" and "men." In Hamilton's words, "Nested within these superordinate categories are lower-order subcategories (possibly at several levels), which are more functional because of their greater articulation of objects (persons) within a category while still proving meaningful differentiations between different categories" (1981, p. 338). Although some stereotyping may take place at the superordinate level, it is assumed that when the observer has some degree of experience with the group in question—such as would be the case with women and men—more differentiated, subordinate concepts are formed.

As Deaux and colleagues (1985) reviewed, there is a good deal of evidence that specific gender subtypes can be identified. (This was highlighted in, e.g., Green & Ashmore 1998, as I discussed above in Section 2.6.) The level at which subtypes exist in the stereotype hierarchy is an issue open to debate, although basic categories are thought to exist at a level where meaningful distinctions can be made but without having so much variation that useful generalizations across gender categories are impossible.

One interesting related point that Deaux and colleagues raised which is worth mentioning here relates to the differences between subtype concepts of males and females, specifically the notion that our culture may allow more latitude in female roles than it does in male roles. There seems to be much more allowable variation in behavior and personal characteristics for people categorized as "women" in general than for people categorized as "men." As Deaux et al. (1985:165) noted:

> Ratings of female subtypes on component measures . . . showed much greater divergence around the general concept of "woman" than did ratings of male subtypes in comparison to the general concept of "man." Inspection of the standard deviations also showed greater variability in judgment of female subtypes . . . the fact that there may appear to be greater diversity in the female subtypes examined here may in itself be an indication of a difference in the ways that images of males and females are stored.

Deaux et al. (1985:166) continued with the idea that in our culture, concepts relating to notions of "male" and "female" may actually be organized differently within the cognitive system:

> Images of men may be centered around a single concept, one that acknowledges little diversity and perhaps tolerates little deviation. Images of women, in contrast, appear to be segmented more sharply. These differences in degree or differentiation should have implication for storage and retrieval processes. They may also suggest some important differences in the extent to which context can influence the stereotyping process. The relative homogeneity of the male stereotype may be more impervious to situational variations, while categorizations of women may be more heavily influenced by the contexts in which the women are viewed.

One prediction that is related to this notion of variation in the relative latitudes of gender roles for women and men is that gender subtyping will likely vary much more across cultures than will the superordinate category, since the latter is grounded in roles that are shared with all other mammalian species. Unlike the seahorse or the penguin, for example, mammalian fathers cannot give birth to the young of the species or incubate the eggs. And while rubber nipples have been in existence for better than 150 years, this is nowhere near long enough to have changed the predominate role of females in feeding the young across cultures—the wet nurse until recently has had to be female.

What this implies, then, is that the schema/prototype model is simply not informative here in an attempt to best model the representation of stereotypes in cognition. A schema/prototype model is simply too rigidly hierarchical, and cannot capture the complexity of perceptual categorization.

In an exemplar model, on the other hand, these "subtype" effects could be represented/accounted for in a very satisfying way—by saying that they are the result of cross-classification of the same exemplar set by other stereotypes, such as class, ethnicity, situation/role, etc. The biggest advantage of representing stereotypes using an exemplar model is that the model will capture any lawful relationship but will not break down if the variation is not linear.

For example, if the exemplar space for parent versus non-parent cross-cuts the exemplar space for male versus female in a way that cannot be captured by orthogonal features, the distribution of exemplars will still capture the relationships nicely.

The issue that is of interest here, of course, is the ways in which different sets of stereotypes interact in perceptual processing. Adopting a schema/prototype model necessarily forces us to think of stereotypes as hierarchical—or, superordinate versus basic-level versus subordinate categories—whereas adopting an exemplar model allows the cross-classification that can capture the interaction of, e.g., gender with phonological categories.

2.8 SUMMARY

In this chapter, I have discussed background issues related to stereotypes that are important for developing a better understanding of the effects of gender stereotypes in speech processing. I have summarized the notion of human perceivers as cognitive misers, which necessarily entails that perceivers will make robust use of stereotypes as aids to processing incoming information. I have also reviewed the various ways that stereotypes may become activated within the perceptual system to further influence processing, including through visual cues from the face of a speaker.

I have also suggested that the representation of stereotypes in cognition is best handled by an exemplar model approach, which allows for the cross-classification of exemplars based on varying aspects of the input signals—more rigidly hierarchical models cannot do this. Human perceivers categorize complex signals in complex ways, and carry out this categorization in the most cognitively efficient way possible, which can be better represented in an exemplar model approach. And, an exemplar approach can also capture the notion that phonological categories are just as much "stereotypes" as are gender stereotypes, a point on which I will elaborate in the following three chapters.

In the next chapter, I will present concepts, methods, and earlier results from a number of relevant speech perception experiments, all of which indicate that socially-mediated stereotypes truly do play a role in the perceptual processing of speech.

CHAPTER 3

STEREOTYPE EFFECTS IN SPEECH PERCEPTION: CONCEPTS, METHODS, AND EARLY RESULTS

3.0 INTRODUCTION

In the past several chapters I have given background information about gender stereotypes, as well as described what seems to be the most advantageous way of thinking about the integration of speaker stereotypes in speech perception. In this chapter, I will describe several speech perception experiments which suggest that socially-mediated stereotypes do indeed affect speech processing.

The following sections describe the concepts that are important to keep in mind throughout this discussion of stereotype effects in speech perception (§3.1), as well as several speech perception experiments (§3.2) which have shown that socially-mediated stereotypes of various kinds impact speech perception. In §3.2.1, I discuss an experiment in which we examined how the mere suggestion of the gender of the speaker influences the perception of vowels. I also describe Niedzielski's studies (1997, 1999) that looked at how the suggestion of speaker nationality influences vowel categorization. In §3.2.2, I describe an experiment in which we examined gendered effects on fricative perception that were induced by seeing speakers as well as hearing them, and in §3.2.3, I describe a similar experiment in which we examined the perception of vowels.

3.1 CONCEPTS

As I discussed in Chapter 1, a traditional normalization-based approach to speech processing is not adequate for characterizing the types of integration that have been shown to occur in speech perception. Normalization simply will not work for the processing of either speech or gender information. It is instead necessary to think of phonological categories as the same type of conventionalized categories (i.e., stereotypes) as are gender, race, and age categories, for example. Once we do that, if we adopt an exemplar model, we can think of different sets of stereotypes as being invoked from (i.e., resonating with) the same set of exemplars, so that perceiving the category "/ʊ/" in *hood*, for example, means perceiving the word form (i.e., the phonetic context effects) *and* speaker gender simultaneously to focus in on the right set of exemplars to recognize "/ʊ/ spoken by a female speaker in the context of /h_d/, etc."

So, we want a model of speech categories as stereotypes that will integrate the phonological and social stereotypes in a way as different social stereotypes have been shown to be integrated. As we perceive gender and race stereotypes in integrated ways, we also perceive phonological and gender, race, etc., categories in integrated and cross-cutting ways.

3.2 METHODS

In this section, I describe several past speech perception experiments, the results of which offer data in support of the notion that stereotypes, especially gender-based stereotypes, affect speech perception. The experiments that I discuss in §3.2.1 considered the perception of auditory information alone, while the experiments that I discuss in §3.2.2 and §3.2.3 addressed the integration of both auditory and visual information in speech perception.

3.2.1 Stereotype Effects by Suggestion

The first experiments that I describe here employ a kind of "inverted form" of matched guise methodology as described in Chapter 2, but instead of assessing perceived personal characteristics on the basis of perceived language as is done in the matched guise technique, here language was perceived on the basis of assumed personal characteristics.

In the first experiment reviewed here, which is discussed more fully in Johnson et al. (1999), we examined the effects that suggestions about an illusory talker have on identification of gender-ambiguous vowels. This audio-only study involved the identification of vowels between /ʊ/ and /ʌ/ on a synthesized F1 continuum in words that ranged from *hood* to *hud*. One group was instructed to imagine a male speaker, and another group was instructed to imagine a female speaker. We hypothesized that listeners would likely use information about group membership of a talker that activates a group stereotype to help them in perceiving speech, even if that information was merely suggested to them.

The audio tokens for this study were constructed by treating a naturally-produced token of *hud* to LPC resynthesis to first extract the voice source, which was then modified so that the F0 was intermediate between that of a stereotypical male voice and a stereotypical female voice (these two voices were also used for stimuli creation for the experiments that I describe in §3.2.2 and §3.2.3). This voice source was then filtered to create a continuum of tokens from *hood* to *hud*, which sounded as if they were produced by a voice of ambiguous gender. Listeners identified the audio tokens in a two-alternative forced choice (2AFC) paradigm in which they were presented with each token over headphones and were told to determine whether the token was an example of either *hood* or *hud*.

The listeners were divided into two groups, one of which was told that the talker was female and the other of which was told that the talker was male. In this way, we could compare perceptions of the same audio tokens when the only experimental manipulation was suggested gender of the talker, to see if this suggestion had an influence on the final speech percept.

We calculated category boundaries between *hood* and *hud* on the F1 continuum for the identification response data. As we expected, when listeners believed that the speaker was female, the boundary between categories was at a higher frequency than when listeners believed that the speaker was male (because listeners expected that that boundary between vowel categories for a female speaker would be at a higher frequency on the F1 continuum than it would be for a male speaker). We concluded that gender suggestion does indeed have an impact on where listeners perceive the vowel category boundaries to be, and that "at least a part of the boundary shift associated with 'talker normalization' is due to expectations regarding male and female voices that listeners bring to speech perception, quite apart from the sensory information presented in the stimuli" (Johnson et al. 1999:380).

In a similar set of experiments that employed a similar instruction set methodology, Niedzielski (1997, 1999) examined the effects that the suggestion of nationality of the speaker would have on listeners in Detroit, Michigan. Niedzielski studied the perception of white, middle-class residents of Detroit, for whom existed strong stereotypes about the English spoken across the Detroit River in Windsor, Ontario, versus the English they spoke in Detroit, as was identified in prior language attitudes studies.

Niedzielski was especially interested in the folk linguistic status of Canadian Raising ("CR," the raising and fronting of the nucleus in the diphthong /aw/, to something more schwa-like) for the Detroit residents, as well as the extent of the Northern Cities Chain Shift ("NCCS"; see Labov 1994) in the area, which existed as a bit of a paradox: the

Detroit speakers consistently identified CR and NCCS vowels in the speech of the Windsor speakers, but never in their own speech, even though acoustic analysis proved that, indeed, many of the Detroit speakers exhibited the same vowel characteristics as did their Canadian neighbors. As a matter of fact, most Detroit speakers felt that their dialect is equivalent to Standard American English, despite these marked areal features.

This knowledge led Niedzielski to ask how it could be the case "that Detroiters do not notice the CR and the NCCS-vowels in the speech of their fellow Detroiters? Do their beliefs about what Detroit residents *should* sound like have greater influence than the acoustic information that their ears receive? If they believe that a speaker is from an area that is stereotyped for these features, would they be more likely to notice them?" (1999:63-64).

To examine this, Niedzielski employed an instruction set methodology similar to that used in our gender suggestion experiment, having Detroit listeners select diphthong onset (Niedzielski's term) tokens that they thought were appropriate examples for a speaker indicated as being either a "MICHIGAN" speaker or a "CANADIAN" speaker. The crux of the experiment, though, was that all of the speakers were actually from Detroit, but were presented twice to each listener with nationality indicated differently each time.

Niedzielski found that when listeners were told that they were hearing a Canadian talker, they were more likely to select a higher onset token, thus representing their expectation for Canadian speakers to exhibit CR. On the other hand, when listeners were told that they were hearing a Michigan talker, they actually selected onset tokens that were *lower* than what the Detroit speaker used in the experiment had produced. Thus, they effectively "lowered" the CR inherent in the Detroit speakers' speech.

As Niedzielski stated, all listeners "heard *the same Detroit speaker*, and because this was the only difference in the two sets of respondents, any differences in the answers from these two sets must be attributed to the expectations that the respondents had based on these two labels" (1999:64). She concluded that "for words that contain the stereotyped /aw/, listeners 'hear' the stereotyped raised variant if the speaker fits the social description of someone who is expected to raise it—that is, someone from Canada. If, however, the speaker does *not* fit this social description—if the speaker is believed to be from Michigan—then listeners are less likely to 'hear' or notice the raised variant" (p. 69).

These experiments show that the suggestion of certain socially-mediated characteristics of the speaker affects speech processing. In the next sections, I describe a number of experiments in which we have examined the integrated processing of speaker categories and phonological categories when information about the speakers comes from other modalities, namely visual.

3.2.2 Gender Effects in Fricative Perception

It has long been accepted in the world of speech perception research that the aspects of speech which contribute to an actual language percept extend far beyond the bounds of acoustic information alone. In addition to acoustic information, visual information about the production of the speech signal is also a contributing factor to perception, and has enjoyed the attention of many researchers over an extended period of time. For more than a century now, for example, researchers interested in improving perceptual ability in noise have studied how it is that watching a speaking face aids in speech perception. This line of research has also been applied to improving perception in people with hearing impairments (see Summerfield 1987 for an earlier review).

Visual information, in fact, has come to be considered a very influential factor in the speech perception process. As the McGurk effect (McGurk & MacDonald 1976, McDonald & McGurk 1978) originally demonstrated, even when the visual signal is mismatched for the place of articulation of the acoustic signal (e.g., when visual /ga/ is synchronized with audio /ba/ in a movie of a talking mouth), the visual information about place is integrated with the acoustic information to give a final percept that is a combination of the two signals—the fused percept /da/ tends to results from the synchronization of visual /ga/ and audio /ba/. Such audiovisual integration in speech perception is repeatedly shown to be a consistent, robust effect for most perceivers.[1] (See Johnson et al. (1999) for a review of McGurk-type effects found in both consonant place and vowel perception.)

But recent evidence indicates that there is more than merely phonetic information about place of articulation in the visual signal that is incorporated into the final speech percept. Strand and Johnson (1996) first described evidence to suggest that the perception of the gender of a speaker is also incorporated into the final speech percept. We conducted two experiments to examine the effects that varying gender percepts have on the identification of synthetic fricatives ranging on a frequency continuum between /ʃ/ and /s/. (The synthetic fricatives were concatenated with naturally-produced vowel-consonant (-VC) segments to produce stimuli that sounded like *shod* or *sod*.)

The first experiment involved the identification of audio-only tokens of monosyllabic fricative-onset words in a 2AFC paradigm, and the second experiment involved the identification of audiovisual tokens that were constructed by dubbing these original audio-only monosyllabic words with movies of synchronized talking heads (again,

[1] Although a number of researchers have shown that across cultures, the McGurk effect occurs to varying extents, such that, e.g., Japanese speakers show a much lower percentage of McGurk integrations than do American English speakers. (See Sekiyama & Tohkura 1991; see also Sekiyama & Tohkura 1993; Sekiyama, Braida, Nishino, Hayashi, & Tuyo 1995; and Sekiyama 1994. For research on the McGurk effect in other languages, see, e.g., Sekiyama 1997 [Chinese], Grassegger 1995 [German, Hungarian], and Fuster-Duran 1996 [Spanish, German]). Researchers have also noted that, for reasons which are not yet fully understood, some individuals regardless of cultural membership are impervious to the McGurk effect (e.g., Munhall, Gribble, Sacco, & Ward 1996:355). Anecdotally, two close aquaintances of mine, both of whom hold Ph.D.s in linguistics, report experiencing no McGurk integration when they view my McGurk demonstration video tape.

2AFC). Both stereotypical and nonstereotypical natural female and male voices were employed in creating the audio -VC portions of the tokens, and female and male faces were employed as talking heads in creating the audiovisual tokens.[2]

We analyzed the shifts in identification functions for the onset fricative noises, and found that in the audio-only condition, listeners' assumptions about the gender of the speaker, as conveyed through aspects of voice quality, significantly affected the way that they categorized the synthetic fricative onset portion of the audio stimuli. When listeners judged the voice producing the coda portion of the token to be stereotypically male, for example, they identified the boundary between synthetic /ʃ/ and /s/ to be at a lower frequency than when they judged the voice producing the same portion of the token to be stereotypically female.

Perhaps more interestingly, there were significant shifts in boundary locations even within gender groups: the nonstereotypical male voice was perceived as producing a category shift at a higher frequency than the stereotypical male voice, and likewise, the nonstereotypical female voice was perceived as producing a category shift at a lower frequency than the stereotypical female voice. Listeners were using information about the natural voices they heard to "back-perceive" the categories to which the token-onset synthetic fricatives belonged.

The second component of these experiments involved the identification of audiovisual tokens, and is perhaps more interesting for the purposes of the present work. In this part of the study, the same synthetic fricative-onset monosyllables were

---

2 Stereotypicality was assessed by first having a separate set of participants rate a group of faces and voices on a 5-point unidimensional scale from 1 = "most masculine" to 5 = "most feminine." Judgment scores were averaged across participants. We then selected the face and voice with an average rating closest to 1 as the most stereotypically male, and the face and voice with an average rating closest to 5 as most stereotypically female. Faces and voices with average ratings closest to 3 were selected as most nonstereotypically male and female for their respective gender categories. (There were no male faces nor voices with average ratings higher than 3, and no female faces nor voices with average ratings lower than 3.)

synchronized with video clips of faces saying the same words. Voices and faces were both mixed and matched for gender, such that each voice was dubbed with several different talking faces of varying gender. Our prediction was that if perceivers were indeed incorporating visual information from the face into their perception of the acoustic signal, we would see evidence of this from shifts in identification functions of the onset fricatives when the audio portion of the token was held constant but the visual portion was varied.

Analysis of the identification functions supported our hypothesis—an audiovisual token with a male face induced a shift in fricative categorization to occur at a lower frequency than if the face in the audiovisual token was female (holding the acoustic portion of the token constant between both audiovisual conditions). Again, we saw shifts that were congruent with what we had found in the audio-only condition, except here, in the audiovisual condition, the shifts were wholly influenced by the faces in the tokens.

We concluded, as I further elaborated in Strand (1999), that from the faces of the talkers in the audiovisual tokens, perceivers formed certain expectations about what the speech of these talkers should be like, and these expectations (or stereotypes) were integrated into the final speech percept. So, our results support the notion that gender stereotypes affect speech perception.

While many current speech perception theories can accommodate the influence of visual phonetic information about place of articulation, such as was examined in the McGurk effect, most theories can't handle this type of visual normalization effect, in which the visual input into perception is linked to socially-mediated expectations.

3.2.3 Gender Effects in Vowel Perception

The following study, more fully described in Johnson et al. (1999), again had two parts. This study followed the methodology employed by the second experiment described above (and first in Strand & Johnson 1996) to show that similar gender effects occur in audiovisual integration in vowel perception, in this case for the vowels between /ʌ/ and /ʊ/ on a synthesized F1 continuum. The first component of the study employed audiovisual tokens for which the audio portions were completely synthesized words varying on a continuum between *hood* and *hud*, and the visual portions were movies of talking heads saying *hud*.

The audio portions of the stimuli were created by combining voice source parameters extracted from stereotypical and nonstereotypical male and female voices (again, four voices total) with the prespecified vocal tract filter parameters that varied systematically across the stimuli continua. Each of the four resulting source functions were then filtered to create a continuum of stimuli ranging from *hood* to *hud*. As Johnson et al. (1999:368) stated, "the four continua produced by this method thus had identical formant trajectories and only differed in voice source." In creating the final audiovisual tokens, voices in the audio portions and faces in the visual portions were again synchronized, and mixed and matched for gender.

Perceivers were again presented with the audiovisual tokens for identification in a 2AFC paradigm. The resulting identification data were analyzed by employing linear interpolation to determine the perceivers' category boundaries based on F1, which were then subjected to repeated-measures analyses of variance, with face gender, voice gender, and voice stereotypicality as factors for consideration. Results indicate similar effects to what we had found in our audiovisual fricative experiment: in terms of the face effect, for stimuli with a female talking face, the boundary between *hood* and *hud* was perceived to be at a higher frequency than it was for stimuli with a male talking face. This follows our prediction that perceivers modify their perceptions of speech based on expectations about how that speech should sound, as drawn from the faces of speakers.

In terms of the voice effect, the boundary between *hood* and *hud* was perceived to be at a higher F1 frequency for female speakers than it was for male speakers. This, again, taps expectations that listeners have about how males' and females' voices "should" sound, rather than relying on acoustic information alone. Finally, the factors of voice stereotypicality and voice gender were found to interact. As Johnson et al. (1999:370) explained this result, "the lowest boundary on the F1 continuum occurred with the stereotypical male voice (564 Hz) and the highest F1 boundary occurred with the stereotypical female voice (617 Hz). The two non-stereotypical voices had boundaries between these extremes (non-stereotypical male at 586 Hz, and non-stereotypical female at 603 Hz)."

Johnson and colleagues pointed out that one potential problem with this first audiovisual vowel perception experiment is that the visual syllables used in token construction were always productions of the word *hud*. This made it impossible to determine whether gender information from the face is integrated into speech perception in the same manner as is phonetic information from the face. The second component in our set of vowel perception experiments addressed this issue.

The second experiment followed the same methodology and means of analysis as did the first experiment, except that this time movies of talking faces saying both *hood* and *hud* were dubbed onto the audio portions from the first experiment. This allowed us to test the McGurk effect per se using our own stimuli. After presenting the stimuli to perceivers for identification and analyzing the data, we found that the word produced by the talking face did indeed have a significant effect on perceivers' indentification of the boundary between *hood* and *hud* on the F1 continuum. That is, when perceivers saw a face saying *hood*, listeners were more likely to label the stimulus "hood."

Face gender was again a significant factor. When perceivers saw a female face, they perceived the category boundary at a higher F1 frequency than when they saw a male face—again, this is logical, considering that perceivers are conditioned to expect that female speakers' formants will be at higher frequencies. The voice gender effect was also significant in this second experiment, but interestingly, it was larger for the stereotypical voices than for the nonstereotypical voices. Also, the difference between the two female voices was greater than it was for the male voices.

This second experiment points again to cross-cutting female versus male stereotypes, as well as /ʌ/ versus /ʊ/ stereotypes—that is, it is impossible to separate them out by "level" or even by modality.

One final important point has to do with the voice quality of the female voices. In our fricative experiment described above, the F0 of the prototypical male voice was lowest and the F0 of the prototypical female voice was highest. The nonprototypical male and female voices fell in the middle. This made it possible to claim that perhaps the voice effect was not so much a gender effect per se, but rather the result of indexical vocal tract length normalization. While this account fit the previous data, the stimuli here give interesting support to the effect being induced, instead, by gender expectations, as we had originally claimed: the prototypically-judged female voice in our vowel experiment actually had a *lower* F0 than did the nonprototypically-judged voice, which makes the vocal tract normalization account untenable.

We can see, however, that the level of breathy voicing is greater in the female voice judged to be stereotypical, which is evident in a steeper spectral tilt. So, listeners are using more than just vocal tract length normalization in perception of these tokens, but are relying also on expectations drawn from other aspects of voice quality. (Henton & Bladon 1985 addressed breathy voice as a marker of female speakers of English, but it must be emphasized that this marker is very culturally-specific. For example, Italian women tend to have more pressed voice than breathy—a quality that in American culture invokes a particular female subtype, the "sultry" Talulah Bankhead voice. Breathiness as an aspect of voice gender is clearly a cultural product and not one of physiology alone, since puberty does not change phonation type habits necessarily in the way that it does vocal tract length.)

As Johnson et al. (1999) concluded from the supporting evidence described in the experiments in the sections above, "talker normalization in speech perception is based on abstract, subjective talker representations [and] listeners perceive talker identity from the totality of information available in the listening situation—including direct acoustic cues for vocal tract length (formants), indirect cues such as F0 and mode of vocal fold phonation [the breathy voice of the stereotypical female in our vowel experiment], visual cues, and even imagined talker characteristics." It is clear, then, that listeners really do employ more information than the physical speech signal alone in the perception of speech, but rather rely on a rich set of direct input as well as expectations about how such input "should" interact in processing.

Returning once more to the notion of representation, one nice point about the exemplar model is that this interaction is automatically built in to the representation by the common links to the same exemplars.

3.3 The Face Gender and Gender Suggestion Effects as Stereotype Consequences

The experiments reviewed above in the first sections of this chapter indicate that gender stereotypes influence speech perception. In this section, I further discuss these findings based on the stereotype review in Chapter 2.

As I described earlier, gender schemas (or stereotypes) are assumed to be activated early in the processing of incoming stimuli through the preliminary categorization of human targets as either male or female, based on a variety of gender cues. Many stereotype researchers posit that more fine-grained gender subtypes are actually activated at this point; it seems reasonable, however, to simply say that some level of gender categorization is made via cues. The data suggest that this categorization then further affects processing. Cues to stereotype activation include conscious thought about a target, the mention of a target, or direct interaction with a target. In our audiovisual experiments, cues for gender identity were conveyed via voice and face information, and in our audio-only suggestion set experiment, cues were inferred via the suggestion of sex as well as through the voice. The content of the activated gender schemas direct further processing through the priming of schema-appropriate categories related to the incoming speech signal. The final speech percept, therefore, is actually a transformed representation of the contents of the original acoustic signal, but transformed in a way that extends a good deal beyond traditional normalization routines. Traditional models of normalization do not account for effects such as these stereotype results, but rather attempt to explain them through what could be characterized as either iconic or indexical models.

As I addressed earlier, the current normalization literature seems to suggest two possible models for the incorporation of gender-related information into speech processing. These models address, in effect, the relationship between cues and effects in speech perception, and theorize a process that strips individual speaker information (presumably of all types) from the signal in order to arrive at an underlying, canonical linguistic representation. One such model could be thought of as an *iconic model* of gender/voice differences, in which there is merely a quasi-direct linkage between cue and effect. In this type of model, a single aspect of the signal is assumed to drive the entire effect. Here, a lower F0 iconically signals, for example, lower F2. This model would assume a simple, direct iconic relationship between F0 and sex/gender.

The second model, which could be considered to be an *indexical model*, allows for cues to correlate with a number of effects. In this type of model, formant values, for example, index (or correlate with) vocal tract length, and F0 values index larynx size, etc. So in this case, the gender effect is an epiphenomenon of the perceived physical size that is signaled.

But neither of these types of models allows for the impact of speaker-specific, stereotype-induced information on speech perception, as found in our results. The interaction between the face and voice information here is very low-level, which cannot be characterized by a prototype theory that relies on a hierarchy of stereotypes and subtypes. In an exemplar approach, however, categories connect to the exemplar space such that each exemplar is then cross-classified for each of several syntagmatic category sets so that male versus female, e.g., is a paradigmatic contrast in the slot [gender: ____] in the syntagmatic set { [gender: ____]; [vowel: ____]; etc. }.

The face gender effect, thus, could be considered to be the natural reflection of a symbolic signaling system, whereby any practical salient aspect of the category accesses the appropriate gender stereotype and evokes all of the stereotypically-associated attributes to prime the interpretation of incoming speech information. In a symbolic signalling system, there is not always a direct relationship between all of the cues to gender and the information contained within the signal. As I have discussed above, in many instances the salient cues for speaker gender are not even contained within the speech signal; when the cue is non-existent as such, the relationship between cue and effect is entirely symbolic.

So what do these results imply in terms of viable speech perception models? Essentially, these notions complicate matters quite a bit. As Nygaard and colleagues have reported (e.g., Nygaard, Sommers, & Pisoni 1994; Nygaard & Pisoni 1998), there is a good deal of evidence that suggests that speaker-specific indexical information is not quickly normalized away during perception, but is instead retained in memory and used to facilitate perception of linguistic events. In this view, representations of linguistic events in long-term memory are richly detailed, rather than being merely abstract prototypical categories. Linked with this notion is the idea that perceptual learning of talker identity (in this case, specific voices) aids in the perception of linguistic information from speakers. Familiar voices seem to be more easy to perceive than unfamiliar voices.

Extending these ideas to the gender effects discussed above, I hypothesize that at least for speakers for whom the listener has yet to develop a richly detailed representation, gender stereotypes will be extremely useful aids in the perception of linguistic information. Because perception is the very complex task that it is, with limited signal invariance to rely upon in the process of linguistic categorization, we as perceivers constantly seek "assistance" in performing this task. Since we do not perceive speech in a social vacuum, our perceptual systems have been honed to use any clues that can possibly be extracted about speaker identity to maximize the outcome from the perceptual process while limiting cognitive cost (i.e., we are cognitive misers). Gender stereotypes invoke information from memory that is used to make the process of speech perception more efficient, and the effects of beliefs and expectations about gender really cannot be separated from the linguistic categories themselves.

Returning to the points made by Nygaard and Pisoni (1998) regarding the consequences of perceptual learning, we can still identify an element of naïveté in terms of the assumed nature of symbolic processing:

> . . . linguistic processing and the perception of talker identity are linked in a contingent fashion. Not only is talker information retained along with lexical information, but these two dimensions do not appear to be separable or independent in perception or attention . . . shared, detailed representations take linguistic representations out of the domain of abstract, symbolic units and into the domain of representation and memory for natural events and specific instances of these events.

But, for "the symbolic species," these so-called "abstract symbolic units" (i.e., stereotypes) *are* natural events (see Deacon 1997). The real issue, then, is how symbolic behavior and the symbolic categories on which it is based can emerge as natural consequences of processing episodic memories in a "grammar" of paradigmatic/syntagmatic relationships that allows a "fast mapping" from instances of new types into the exemplar space. In their above quote, however, Nygaard and Pisoni (1998) conceived of symbolic processing as necessarily "transformational" (memories of one thing are transformed into memories of another, less detailed thing), whereas an exemplar model would give us a way of conceiving of symbolic processing as essentially an *organizational* thing (memories of paradigmatically or syntagmatically related things are linked in a highly efficient organization in memory, in a way that lets us go quickly beyond correlation—i.e., merely indexical organization—to invoke the syntagmatic and by that the absent other, the paradigmatically contrasting type that could have occurred in the same slot.)

I contend that stereotypes about gender (and race, and age, and occupation, etc.) are an important component of a percept of talker identity, and will likewise affect encoding and memory processes related to language perception.

3.4 SUMMARY

> . . . gender is not just a variable that influences the "basic" processes of thinking, feeling, appreciating, striving, and acting, but also a social fact that conditions and governs these very processes.
>
> [Cross & Markus 1993:55-56]

In this chapter, I have reviewed a number of speech perception experiments to support my claim that the influence of gender stereotypes on cognitive processing, specifically the perception of speech, is pervasive, even to the point of altering perceptual reality. This being the case, it is imperative that our accounts of speech perception allow for the role of such an influence, even if it complicates our conceptions of how such processing might work.

One potentially confounding factor in our speech perception results is that our assessments of gender stereotypicality for the faces and voices that we used in stimuli construction were made explicitly. That is, we asked participants to judge faces and voices on a unidimensional scale of masculinity/femininity. But as recent work on the implicit nature of stereotype activation and application suggests, this explicit measure may not be the most accurate way of assessing gender stereotypicality.

In Chapter 4, we employ several methods in a novel approach to access participants' *implicit* gender stereotypes about voices and faces.

CHAPTER 4

STEREOTYPICAL SPEAKERS: MULTIDIMENSIONAL SCALING AND GENDER IDENTIFICATION RESULTS

4.0 INTRODUCTION

In the preceding chapter, I discussed the results of research which suggests that socially-mediated stereotypes have an influence on speech perception processing. In this chapter, I focus on the subjective perception of a set of faces and voices, with the intent of selecting a number of gender-stereotypic and gender-nonstereotypic faces and voices for use in stimuli creation for further speech perception experiments. My goal in this chapter, as well as in Chapter 5, is to construct a means for determining whether there is a processing deficit in speech perception due to the impact of socially-mediated gender expectations. That is, do gender stereotypes at a fine-grained level have a role in speech perception? My strategy will be to construct experimental tokens by employing a number of faces and voices that differ on gender stereotypicality measures (which will be determined in the present chapter), and to then measure the impact that these various faces and voices have on perceptual processing (the methods and results for which are described in the following Chapter 5).

There have been a number of approaches taken by past researchers to assess faces and voices on various gender dimensions. Gaudio (1994), for example, used a set of salient semantic differential scales representing stereotypes about gay American English speakers to assess the extent to which various voices sounded "gay" to a set of listeners. The four differential adjective pairs employed by Gaudio (p. 45) were *straight/gay*, *effeminate/masculine*, *reserved/emotional*, and *affected/ordinary*. Grammer and Thornhill (1994), for example, employed 7-point rating scales for salient adjectives in their effort to assess the attractiveness of faces as judged by German viewers. The four adjectives employed by Grammer and Thornhill (p. 233) were *attractive, dominant, sexy*, and *healthy*. In our own past work, as I described in Chapter 3, we used a 5-point unidimensional scale from "most masculine" to "most feminine" to assess judged gender stereotypicality in faces and voices.

The inherent problem with methods such as these, though, is that while they may give an accurate *explicit* measure of some feature or belief, what is really of most interest is an *implicit* measure of the same thing. The comparison can be made to the reliability of a linguist asking a language consultant whether he or she would say something in a certain way or use a particular phrase in a certain context. While the self-reported language use of the consultant might be of interest for a variety of reasons, it most likely does not exactly correlate with actual linguistic performance. In the same way, self-reported (explicit) perceptions of gender stereotypicality of faces and voices may not necessarily correspond exactly to the actual gender stereotypes that are implicit and operative within the cognitive system.

A good deal of recent social psychological research has indicated that gender stereotyping occurs implicitly to affect further cognitive processing, "without subjects' conscious awareness of the source or use of the stereotypic information in judgment" (Banaji & Hardin 1996; see also Banaji & Greenwald 1995; Banaji, Hardin, & Rothman 1993; and Devine 1989). In this vein, it would be useful for our purposes to develop a means of assessing the gender stereotypicality of faces and voices in an *implicit* manner, rather than the explicit manner that we employed in the past when we used our 5-point masculinity–femininity scale. By doing this, we will be better able to assess the potency of gender stereotypes in speech perception, as I discuss in Chapter 5.

The methods that we employed in our attempt to eliminate conscious judgment in face and voice assessment included the assessment of face pair similarity and voice pair similarity, as well as the speeded classification of face and voice gender. The assessment of face and voice similarity involve the use of the statistical technique of multidimensional scaling, which we employed as a novel means of determining stereotypicality without having participants judge it overtly. We used speeded gender classification as a way to "check" our multidimensional scaling results, based on the notion of stereotypes as cognitive shortcuts, as I described in Chapter 2. That is, stereotypes induce faster processing, so faces and voices that are more stereotypical should be categorized more quickly than those that are nonstereotypical.

Each of the four sections of this chapter describes the methods and results of a distinct experiment designed to acquire knowledge about the processing of gender information in both voices and faces. In Experiment 1, we used multidimensional scaling to examine the perception of a set of voices in terms of their similarity to each other. The related Experiment 2 used the same methods to examine the perception of faces. In Experiment 3, we assessed reaction times to various voices using a speeded gender classification task. Experiment 4 used a similar speeded gender classification technique to assess reaction time to a set of faces. In §4.5, then, the results of these four experiments are taken into consideration to select four voices and four faces (from among the stimuli for Experiments 1 through 4), to be used in the creation of further stimuli for the speech perception experiments that I describe in Chapter 5.

4.1 EXPERIMENT 1: MDS FOR VOICE SIMILARITY

4.1.1 Experiment 1 Purpose

This experiment employs the statistical technique of multidimensional scaling (MDS) to assess the similarity between a number of voices. The results from this experiment will be used in conjunction with the reaction time results for the gender identification of voices task (the results of which are described in §4.3) as a means for determining the level of stereotypicality in voices. Based on the results of these two experiments, we will identify a female and a male voice that are both stereotypical, and a male and a female voice that are both nonstereotypical, which we will then use to construct stimuli for further experiments that examine the interaction of face and voice stereotypes and speech processing.

MDS allows the calculation of psychological similarity spaces for items of various types, and in this case is a way to directly tap the perceived similarity of faces, and perceived similarity of voices without explicitly asking subjects their opinions. Thus, it contrasts with our prior use of the 5-point masculinity–femininity scale, in experiments where participants were overtly asked to rate a gender characteristic. That is, the current method gives an implicit measure of perceived similarity, and avoids self-report bias in subjects, as well as bias on our own part in selecting voices and faces to use as stimuli in the following experiments.

4.1.2 Multidimensional Scaling

Multidimensional scaling is set of statistical techniques that are useful for "uncover[ing] the 'hidden structure' of data bases" (Kruskal & Wish 1978:5). Multidimensional scaling can be used to distill meaningful organization from complex sets of data of all kinds, and has applications to any research area for which there is interest in determining the perceived similarity between entities of various types. Linguists have employed MDS, for example, in an effort to quantify the perceived similarity between intonational categories used by a set of labelers trained in the same prosodic transcriptional system (Herman 1998, Herman & McGory 1999), and psychologists have used MDS, for example, as a means of assessing the perceived cultural similarity among the world's nations (Kruskal & Wish 1978:45 cited Wish 1970 and Wish & Carroll 1974).

The input to a multidimensional scaling solution are similarity ratings between all pairs of objects or categories presented to subjects in a rating task, and the output is a multidimensional "map" that reduces the many paired ratings to a more manageable number of dimensions. Stimuli that are judged to be more similar to each other appear closer together in the map space, and stimuli that are judged to be less similar to each other appear further apart in the map space. The number of dimensions in the map space that are needed to represent the relationships between all of the compared objects depends on the number of factors perceived as salient for distinguishing between objects in the set. An object's location within the map space is indicated with a set of Cartesian coordinates, for which the axes radiate from the absolute center of the map space; the number of coordinates necessary to distinguish the object's location in the space will depend on the number of salient dimensions in the MDS solution. The map that results from an MDS treatment of a set of data is essentially a representation of the psychological relationships between all of the points in the data set, as held in the minds of the perceivers.

This concept can be more clearly described with an example using several American cities and their orientation to each other. Considering the east coast cites of Boston, New York City, and Washington, DC, most people who are familiar with the geography of this area know that these three cities are approximately equally spaced on a line running northeast to southwest. If these geographically-savvy people were asked to rate the distance between pairs of the cities using an arbitrary unit, they might rate the Boston–New York City pair as having a distance of 1 unit between them, the New York

City–Washington, DC, pair as having a distance of 1 unit between them as well, and the Boston–Washington, DC, pair as having a distance of 2 units between them. With these distance ratings as input to a multidimensional scaling solution, the result would be a one-dimensional solution with a single x-axis value for each city representing their distances from each other, as shown below in Figure 4.1.

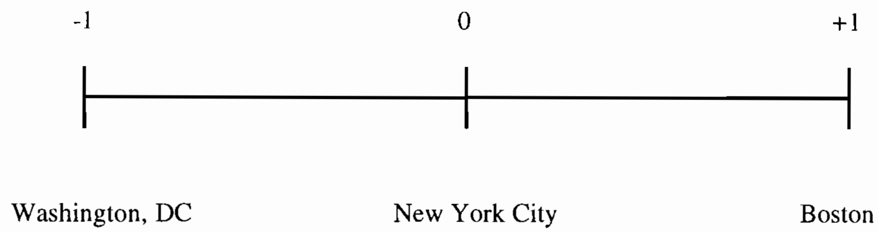

Figure 4.1: One-dimesional scaling solution for the distances between Boston, New York City, and Washington, DC.

But if the city of Pittsburgh, for example, were added to the set of cities for which pairwise distances were rated, then a one-dimensional solution would no longer suffice, since Pittsburgh lies west of the primary Boston—Washington, DC, axis. A second dimension would result from the multidimensional scaling procedure, which would then give each city a set of x- and y-coordinates, therefore allowing all of the cities' respective distances to be accurately plotted in two-dimensional space.

Additional dimensions will come out of the scaling solutions as various factors become salient in rating the distances between stimuli. In the city distances illustration, for example, a three-dimensional solution could be imagined if each city's elevation were in some way taken into account when rating the distances between pairs. Alternatively, if raters' perception of distance is influenced by travelling time, a third dimension might reflect this, because roads between New York City and Boston tend to be less congested than those between New York City and Washington, DC. Similarly, routes between Boston and Pittsburgh are less direct than those between Pittsburgh and Washington, DC.

Multidimensional scaling has been used in the past by linguists to assess the perceived similarity between voices, as well. Murry and Singh (1980), for example, compared MDS solutions for men's and women's voices, and identified a four-dimensional solution in which perceived pitch/F0 was the most important dimension for rating male voices, but nasality/effort was the most important dimension for rating female voices.[1] Gelfer (1993) used MDS analysis to examine the perceptual discrimination of female voices by two groups of listeners, speech pathologists and untrained listeners. She interpreted a five-dimensional solution as being the most accountable for the speech pathologists' discrimination ratings, but only a two-dimensional solution for the group of untrained listeners, with perceived pitch being the primary dimension for both groups.[2]

As speech pathologists are naturally more accustomed to isolating and identifying various voice qualities than are untrained listeners, it is not surprising that the group of speech pathologists showed a greater number of interpetable dimensions in their judgment of the stimuli voices. This also illustrates the point that the number of dimensions that fall out from a multidimensional scaling analysis is directly related to the factors used for judgment that are salient in the minds of those performing the judgment task.

---

[1] The four dimensions that were reported in Murry & Singh (1980) included nasality/effort, perceived pitch, and duration. They were unable to interpret the fourth dimension.

[2] The five dimensions that Gelfer (1993) interpreted for the listening group comprised of speech pathologists included perceived pitch, perceived loudness, perceived age and speech rate (with slower speakers perceived as being older), variability in perceived pitch, and perceived voice quality. The two dimensions that were interpreted for the untrained listeners were perceived pitch and a resonant-shrill quality (combined in the first dimension), and perceptual judgments of pitch variability, perceived age, and speech rate (combined in the second dimension). Gelfer used the listeners' ratings of each voice on several semantic differential scales to help her interpret the dimensions in the MDS solutions.

We employed MDS techniques for assessing voice similarity here as a means of learning more about the specific gender-associated characteristics of the voices that we will to use in further experimentation, specifically the voices of the 20 speakers as described in §4.1.3.1. Most importantly for our purposes of assessing *implicit* measures of gender, we did not tell participants which gender characteristics should be considered in their assessments of similarity; rather, those characteristics that were salient for the participants drove their judgments. These characteristics will, hopefully, be uncovered as the dimensions in the MDS solution. We then used the resulting similarity ratings as one of a number of implicit measures to help us identify the most typical and least typical voices in our set. When combined with the voice reaction time results and gender judgment accuracy results that are described in §4.3, we were able to distill measures of voice stereotypicality for our voices. We then used these stereotypicality measures to create the stimuli for the speech perception experiments that I describe in Chapter 5.

4.1.3 Experiment 1 Methods

4.1.3.1 Participants

The participants for Experiment 1 were 24 undergraduate and graduate students at The Ohio State University (15 females, 9 males; 13 undergraduates, 11 graduates). An additional 20 students (10 males, 10 females; 5 graduates, 15 undergraduates) served as speakers to create the stimuli for the experiment, as described below in §4.1.3.2. Some of the participants received partial course credit in their undergraduate linguistics courses in exchange for their participation, with the remaining participants receiving $5.00 for their participation. All were native speakers of American English, and all completed personal history questionnaires before taking part in the experimental session. None reported any speech or hearing disorders. Each of the experimental participants took part in both Experiment 1 and Experiment 2, and the duration of the complete experimental session was approximately one hour. All participants were debriefed immediately after their session.

4.1.3.2 Stimuli

Stimuli for Experiment 1 consisted of 480 digitized sound files created from audio tape recordings of 24 naturally-produced English words, each spoken by 20 speakers (10 males and 10 females) who were undergraduate and graduate students at The Ohio State University. To minimize potential lexical and segmental effects, members of the set of words were phonetically-balanced (see Egan 1949), and were matched for frequency, word familiarity, and lexical density, as indicated in the on-line Hoosier Mental Lexicon database (which is based on Webster's Pocket Dictionary; see Nusbaum, Pisoni, & Davis 1984). See Appendix A for a list of these 24 words.

Audio tape recordings were made of each of the 20 speakers uttering these word lists for creation of the audio stimuli used in Experiment 1. (The speakers were simultaneously video recorded for creation of the stimuli used in Experiment 2; see §4.2.3.2 for description of the video recording procedure.) The recording session took place in a quiet room. Each speaker sat in a chair facing a video camera, directly above which the word lists were displayed. A Shure SM2 Dynamic Cardioid microphone was mounted on a table stand and placed approximately 2 feet from the speaker's mouth, at a 45 degree angle to the speaker's left side, just out of view of the video camera range. A Symetrix SX202 Dual Mic Preamp was used to preamplify the speech signal to the proper level, and a Denon DRM-700 Stereo Cassette Tape Deck was used to make the recordings (Dolby Noise Reduction set to "Off").

The tape-recorded utterances of these 24 words as spoken by the 20 speakers were first transformed into digitized computer sound files using Macromedia SoundEdit 16 software, version 2, on a Power Macintosh G3 computer. The tape-recorded utterances were preamplified using an Audio-Technica ATUS AM100 Stereo Mixer, and were digitized at a sampling rate of 22.05 kHz. The SoundEdit 16 software was further used to edit the contents of each sound file by cropping the file so that it started at the onset of waveform energy for the first speech sound of the word, and ended at the closure or release of the final speech sound in the word.

So in terms of onset, for example, the sound file for a fricative-initial word such as *sheet* was cropped to begin at the onset of fricative energy, while the sound file for a voiceless stop-intial word such as *town* was cropped to begin at the stop release burst of the initial [t].

Word-final speech sounds were treated in a similar fashion; it should be noted, however, that there is a good amount of individual and dialectal variation in terms of whether word-final stops in words such as *sheet* are released or unreleased, which is reflected in these stimuli. Figure 4.2 displays a sample waveform and spectrogram which correspond to a single edited sound file, as originally produced by Male Speaker 8 (the final consonant sound in this utterance of *sheet* is unreleased by this speaker).

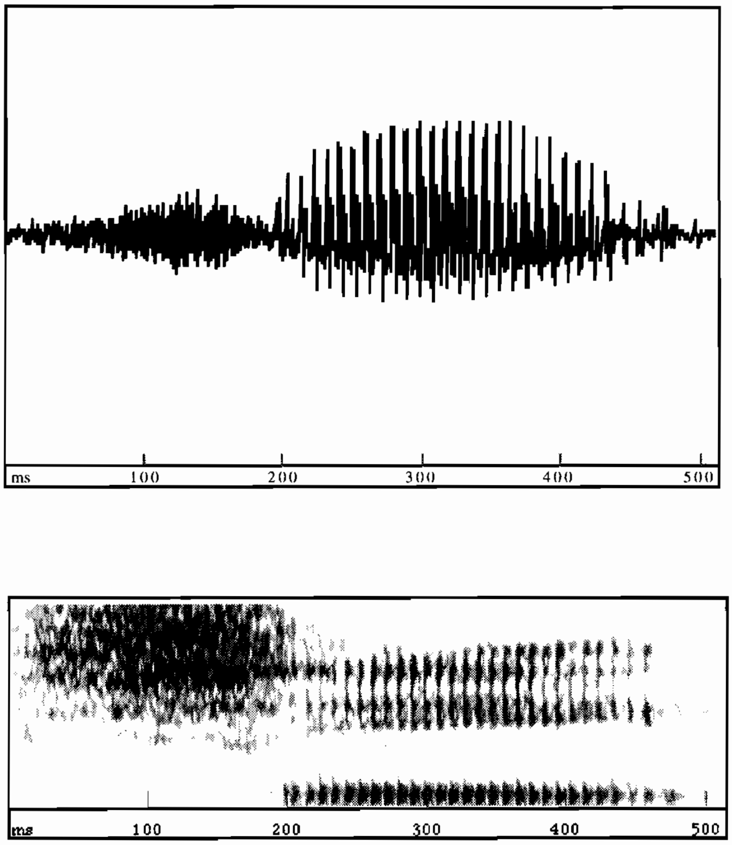

Figure 4.2: Waveform (top panel) and spectrogram (bottom panel) representing the sound file for the stimulus word *sheet*, as originally produced by Male Speaker 8. The onset of the sound file corresponds with the onset of initial fricative noise, and in this case ends at the onset of the final closure. (The horizontal axis in each display indicates time in milliseconds. The vertical axis in the top waveform panel indicates the amplitude of air pressure during the utterance, relative to the surrounding air pressure. The vertical axis in the bottom spectrogram panel indicates frequency in herz (Hz), ranging from 0 to 5000 Hz, while darkness indicates the relative intensity of energy at that frequency.)

The waveforms of each of the 480 sound files were scaled so that the peak root-mean-square (RMS) amplitude values were equated for all files, at an amplitude of approximately 75 dB. (*RMS amplitude* is a measure of acoustic intensity that is closely related to perceived loudness; see Johnson 1997:36-37 for further description of RMS amplitude and its calculation.)

While there was variation in the racial/ethnic background of the speakers, all were native speakers of American English varieties. Of the 10 female speakers, 6 characterized themselves as Caucasian/European-American/white, 1 characterized herself as Korean-American, 1 as Asian-American, 1 as Middle Eastern-American, and a final characterized herself as Indian-American (i.e., ancestors from India). Of the 10 male speakers, 7 characterized themselves as Caucasian/white, and 3 characterized themselves as African-American/black.

4.1.3.3 Procedure

Before starting Experiment 1, each participant read a printed list of experiment instructions in which they were told that they would be hearing a series of voice pairs, and that their task would be to judge HOW SIMILAR the voices in each pair sounded according to the 9-point scale illustrated below in Figure 4.3.

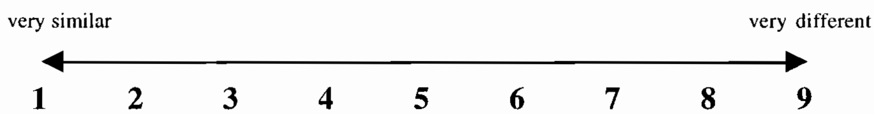

Figure 4.3: The 9-point similarity scale used for judging voice similarity in Experiment 1.

In making their judgments, participants were instructed to focus on each speaker's voice in general rather than on the way that the speaker said the word. They were told that it might be helpful to ask themselves how likely it was that the speakers in the pair could be twins; a judgment of "very likely twins" would rate a "1" on the 9-point scale. The listeners heard each speaker in each pair utter the same word.

Participants sat in a sound-attenuated booth and were presented pairs of the experiment stimuli as produced by different speakers. Each of the 24 participants heard only one of the 24 stimulus words, but each participant heard all pairwise comparisons of the 20 speakers uttering the word, giving each participant a total of 190 distinct pairwise comparisons to be judged for similarity on the 9-point scale.

MEL Professional software on a Dell 466/M computer was used to present the stimuli to participants and collect their responses. Participants heard the stimuli sets binaurally via Beyer Dynamic DT 220 High Fidelity headphones, with the first and second stimulus in each pair separated by a 500 msec interstimulus interval (ISI). Participants entered their similarity judgments using the keys numbered 1 through 9 on the computer keyboard. After each judgment was logged, participants immediately saw a statement on the computer monitor indicating what their similarity judgment had been. Participants were given up to 5 seconds to respond with a similarity judgment, otherwise an indicator of "no response given" was logged. The intertrial interval was 1000 msec, after a response or "no response given" indicator was logged.

Participants first heard a practice block of 5 voice pairs to make them comfortable with the procedure, as well as to give them a sense of the range of voices that would be included in the experiment. Both orders (AB and BA) of each of the 190 pairwise comparisons were then presented in random order to each participant, for a total of 380 pairs for similarity judgment. These 380 pairs were broken into 5 blocks of 76 stimuli pairs each.

4.1.4 Experiment 1 Results and Discussion

After first acquiring the similarity judgments, there are several steps that must be taken to obtain a multidimensional scaling solution. First, the similarity judgments must be put into a symmetric matrix so that judgments for each pairwise comparison are repeated in both the lower-left and upper-right halves. In this way, each row of the matrix is an input dimension indicating perceived similarity between a particular voice and all of the others in the set. The object is now to reduce this 20-dimensional space (20 rows for 20 speakers) to a more tractable set of meaningful dimensions of similarity.

Next, this matrix of similarity judgments must be treated to the multidimensional scaling statistical techniques themselves, with the result being the creation of the multidimensional map space. And finally, the resulting solutions must be interpreted to determine the most accurate number of dimensions necessary to give the best final MDS solution for the data.

This interpretation step requires the consideration of a number of factors, including the R^2 value given for a solution. The R^2 value will range between 0 and 1, and represents the proportion of variance in the data that is accounted for by that specific MDS mapping. The closer the R^2 value is to 1, the better the MDS solution accounts for the data. The general interpretability of the dimensions and their relationship to the patterning of the data is also an important factor for consideration when trying to determine the most advantageous number of dimensions to be included in the MDS solution.

It should be noted that maximal R^2 value, however, is not valued as highly in an MDS solution as are clearly interpretable dimensions. An MDS solution with the fewest dimensions possible is always preferred over one with too many, even if the latter may have a slightly higher R^2 value. As Kruskal and Wish (1978:45) stated, it is most important to "use any means at your disposal to understand as much of the data and results as possible," rather than relying on a single value to determine the best MDS solution.

In the case of our data, we first calculated averages of the similarity judgments for each voice pair across all words and all participants. We then transformed the set of averaged similarity ratings into a symmetrical matrix of data that was appropriate for further MDS analysis. The matrix of similarity judgments for the pairs of voice stimuli that were judged in Experiment 1 is given in Table B.1 in Appendix B.

Next, the matrix of voice similarity judgments was treated to an MDS analysis using Systat, version 5.2.1, software. All MDS calculation options were left at the original default values[3] in the Systat program, with the exception of the loss function, which was changed to "Guttman's coefficient of alienation" from the original "Kruskal's stress formula 1" Systat specification.

The two-dimensional scaling solution for perceived voice similarity that was chosen as the best fit for our data is shown in Figure 4.4. The R^2 value for this solution is .98957, with two clear clusterings of data for the male and female voices. Dimension 1 clearly represents this gender aspect, and is graphed along the horizontal axis. Dimension 2 is graphed along the vertical axis, and while it is not immediately apparent what this dimension represents, it is clear that the data for both male and female voices tend to vary more or less with respect to this second dimension as well.

---

[3] Default MDS values in Systat include Decrement=0, Iterations=50, and Minimum=0; these three options are methods for controlling the number of iterations necessary to produce the best MDS solution. A fourth default value, Minkowski Constant=2, specifies that Euclidean distances should be calculated between points in the MDS space (versus "city block distances," for which the Minkowski Constant would be specified as "1"). The final option default value, Shepard Plots=1, generates a single Shepard diagram to plot distances against similarity values for the final solution. (See Wilkinson, Hill & Vang 1992:101-102 for further description of these Systat options and default values.)

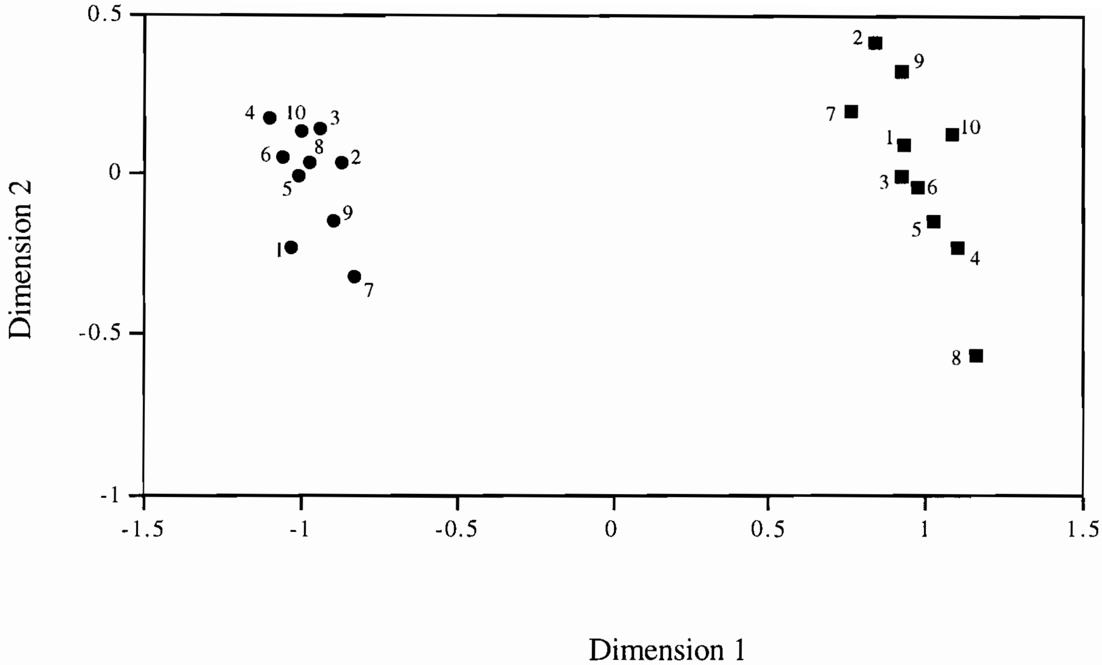

Figure 4.4: Two-dimensional scaling solution for voice similarity of the voice stimuli in Experiment 1. Female voices are indicated by filled circles, and male voices are indicated by filled squares. The numbers that mark each token correspond to the specific speaker for which the MDS value was calculated. Ticks along the axes are labeled with coordinates that correspond to the space of the MDS map, with the tick labeled "0" corresponding to the center of the specific dimension. These values are not absolute in any way, but are relative only to the arrangement of this set of data in this space.

The multidimensional scaling solution that is represented in Figure 4.4 shows a very clear separation between the mapping of the female and male voices as groups, with female voices clumping toward the left of the map space (represented by filled circles), and male voices clumping toward the right of the map space (represented by filled squares); again, Dimension 1 is clearly equivalent to basic gender categorization. This very apparent separation in the map space indicates that participants robustly rated female voices as much more similar to other female voices, and male voices as much more similar to other male voices. While this result may not seem surprising, it is interesting to see such a clear division in light of my subjective impressions of a particular few of the voices in relation to the others: to my ear, Female Speaker 7 sounds much closer to Male Speaker 1, for example, than Female Speaker 7 sounds to Female Speaker 6, yet both female voices appear closer to each other in the perceptual space than either of them does to any of the male voices. Clearly, sex/gender was the primary factor that participants used in rating the similarity of these voices, and it was a very robust primary factor.

Dimension 2 might plausibly be interpreted as representing some voice quality associated with femininity/masculinity. As I will discuss later in this chapter, while all of the male voices were accurately judged to be "male" at near-ceiling levels, there was a good deal of variation in the accuracy of gender judgment for the female voices. The two female voices toward the bottom of the cluster (Female Speaker 1 and Female Speaker 7) showed the lowest accuracy scores for speeded gender judgment, but they still clustered with the other female voices in general in the MDS solution for similarity judgment. This suggests that the variation in accuracy was completely orthogonal to the gender feature that split the male and female voice groups.

Also of interest is the fact that there is more apparent spread for the male voices in this perceptual space than there is for the female voices. While this may be simply an artifact of the variation inherent in the set of voices employed in the similarity judgment task (i.e., the male voices could have been more varied as a group than the female voices, which could then show up in their more disparate mapping), it could also be representative of the differences in perception of males and females in American culture in general: as I mentioned in Chapter 2, females in Western cultures (and many non-Western cultures, for that matter) are generally allowed much more latitude in all aspects of their gender performance, while still being perceived as "good female exemplars." Males, on the other hand, are culturally required to project a much more restricted set of gender characteristics in order to be perceived as "good male exemplars." Therefore, variation among the female voices might be collapsed into a smaller perceptual space since all sound "more or less like (different types of) good women," while variation among male voices might tend to be perceptually more salient to listeners, which could lead to an expanded perceptual space as is apparent in Figure 4.4. (I will return to these issues in §4.5 and §4.6, when I discuss our final selection of voices for use in further stimuli construction, with respect to all of the results regarding voice gender perception in this chapter.)

Before I address the use of multidimensional scaling to assess face similarity in §4.2, it is necessary to say a few words here about what these MDS results might indicate about stereotypicality in general. These experiments are based on the notion that the assessment of perceived similarity as rendered by the MDS solution can be linked to a similar measure of perceived stereotypicality. It has been claimed that faces and voices that are located toward the central regions of their respective gender clusters on the MDS solutions' gender dimensions are likely to also be perceived as more stereotypical than faces and voices that are located away from the center clusters.

In his study of facial similarity spaces, for example, Busey (1998:476) stated that "the most typical face will appear in the center of the space." In their research on attractiveness, Langlois and Roggman (1990) showed that faces judged as attractive are those that represent the central tendancy or averaged members of categories of faces.

Stereotypes aid cognitive processing by serving as shortcuts for avoiding extraneous processing that can be equated with the most typical category members; the supposition exists, therefore, that the most typical category members for the voices represented in Figure 4.4 will be the voices at the centers of the gender clusters. Nonstereotypical voices, likewise, are claimed to be represented as the outliers, away from the centers of the clusters.

Knowing that gender-stereotypic exemplars should be responded to more quickly than non-stereotypic exemplars (e.g., Banaji & Hardin 1996), it should follow that the voices at the center of the MDS spaces here should be responded to more quickly than those that are located further from the center. The speeded voice gender classification task which I describe later in §4.3 assesses this interpretation of the MDS space.

4.2 EXPERIMENT 2: MDS FOR FACE SIMILARITY

4.2.1 Experiment 2 Purpose

This experiment follows the MDS techniques that were used in Experiment 1 to assess the similarity between voices, but in Experiment 2 we assessed the similarity between members of a set of faces. The results from this experiment were used in conjunction with the reaction time results for the gender identification of faces task (the results of which are described in §4.4) as a novel means for determining the level of stereotypicality in faces. Based on the results of these two experiments, we identified female and male faces that are both stereotypical and nonstereotypical, which we used to construct stimuli for further experiments that examine the interaction of face and voice stereotypes and speech processing, as I discuss in Chapter 5.

4.2.2 Multidimensional Scaling for Face Similarity

The use of multidimensional scaling to assess similarity between faces has become a common procedure in the research, for example, on the ways that faces are represented in memory. Perhaps the most generally-accepted model of face memory is Valentine's (1991a, 1991b) exemplar model of multidimensional face space, in which faces are stored as locations in multidimensional space and are organized according to similarity.

Along these lines, face perception researchers have used MDS in an attempt to determine which features perceivers use in assessing facial similarity; each salient feature in an MDS solution, of course, corresponds to a single dimension in the solution. Busey (1998), for example, determined that for the set of male faces that he tested, a 6-dimensional solution could be interpreted, with the dimensions corresponding to age, race, facial fatness, amount of facial hair, long versus squat head shape, and color of facial hair. As I mentioned in the previous section, Busey (1998:476) stated that the most typical face will appear at the center of the face space in an MDS solution, which can be interpreted to represent the most stereotypical. We assume that notion here by having participants rate the similarity of 10 female and 10 male faces.

4.2.3 Experiment 2 Methods

4.2.3.1 Participants

The participants for Experiment 2 were the same 24 Ohio State undergraduate and graduate students who participated in Experiment 1, as discussed above in §4.1.3.1 on page 91.

4.2.3.2 Stimuli

The stimuli for Experiment 2 consisted of 20 still pictures of the faces of 20 different people (10 male, 10 female), taken from videotapes of them reading the word lists from which the stimuli for Experiment 1 were constructed. These speakers were video- and audio-taped in a quiet room, with a neutral backdrop and directional lighting on each side of the speakers' faces, at 45-degree angles from front-orientation. The video camera was zoomed in to close-up range to capture a consistent head-and-shoulders image of each speaker. Seven of the females and six of the males whose faces were used to create the visual stimuli for this experiment also supplied their voices for audio stimuli creation in Experiment 1, but voice stimulus number and face stimulus number do not always correspond.[4]

To create the still pictures of the 20 faces, single frame head shots from videotape were transformed into digitized computer picture files using Adobe Premiere LE software, version 4.2.1, on a Power Macintosh G3 computer. Segments of the video were selected for digitization where the speakers' mouths were closed, their faces and eyes were oriented toward the camera, and their heads were oriented on a relatively vertical axis. GraphicConverter software, version 3.8, was used to maximize the contrast and hue values of the picture files in order to enhance the distinctiveness of facial features, textures, and coloration upon presentation to participants.

The picture files were saved at a display size of 320 x 240 pixels, or about 11.3 cm wide by 8.5 cm high, and were then translated into pcx video file format, which is

[4] Stimulus number correspondences for these talkers are as follows, with the face and voice stimuli that represent the same talker separated here by a dash, and each talker's self-identified race/ethnicity given in parentheses: Female Face 1—Female Speaker 1 (Korean-American); Female Face 4—Female Speaker 4 (White); Female Face 5—Female Speaker 5 (Asian-American); Female Face 6—Female Speaker 6 (Caucasian); Female Face 7—Female Speaker 7 (White); Female Face 8—Female Speaker 8 (Middle Eastern-American); and Female Face 10—Female Speaker 10 (Indian-American). Talker information for the males is as follows: Male Face 1—Male Speaker 1 (White); Male Face 4—Male Speaker 5 (African-American); Male Face 5—Male Speaker 7 (Caucasian); Male Face 6—Male Speaker 8 (African-American); Male Face 7—Male Speaker 9 (Black); Male Face 10—Male Speaker 3 (White).

compatible with the pc-based MEL Professional experiment-running software on which Experiment 2 was run. See Appendix C for a display of the 20 face stimuli used in Experiment 2.

4.2.3.3 Procedure

Before starting Experiment 2, each participant read a printed list of experiment instructions in which they were told that they would be seeing a series of face pairs, and that their task would be to judge HOW SIMILAR the faces in each pair appeared according to the 9-point scale illustrated below in Figure 4.5.

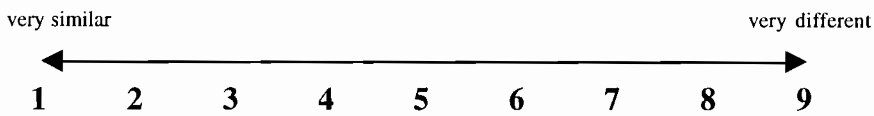

Figure 4.5. The 9-point similarity scale used for judging face similarity in Experiment 2.

In making their judgments, participants were instructed to focus on each speaker's face in general, rather than on details such as the color of their clothes or whether one's face might appear slightly larger than the other's because the level of zoom of the video camera varied slightly from face to face during the original video taping. Participants were told that it might again be helpful to ask themselves how likely it was that the speakers in the pair could be twins; a judgment of "very likely twins" would rate a "1" on the 9-point scale.

Participants sat in the same sound-attenuated booth for Experiment 2 as they had for Experiment 1. MEL Professional software on a Dell 466/M computer was used to present the visual stimuli to participants and collect their responses. They saw the visual stimuli pairs displayed in full color on the computer monitor, presented side-by-side on a black background with approximately 1 inch of black space between the two (the display size of each stimulus, as mentioned above, was approximately 11.3 cm wide by 8.5 cm high). Participants again entered their similarity judgments using the keys numbered 1 through 9 on the computer keyboard. They were given up to 5 seconds to respond with a similarity judgment, otherwise an indicator of "no response given" was logged. After entering a response, an intertrial interval of 1500 msec preceded presentation of the next set of faces for judgment.

4.2.4 Experiment 2 Results and Discussion

The procedure for creating a multidimensional scaling solution that was described earlier in §4.1.4 was followed here to analyze our face similarity judgments. The matrix of similarity judgments for the pairs of faces that was input into the MDS statistical analysis is given in Table D.1 in Appendix D.

The two-dimensional scaling solution for perceived face similarity that was chosen as the best fit for our data is shown below in Figure 4.6. The R^2 value for this solution is .87100, which accounts for a good amount of the variation in the mapping. The three-dimensional solution that was posited for this data showed a slightly increased R^2 value of .92869, but the increased difficulty of interpreting the three-dimensional space led us to select the two-dimensional scaling solution as the optimal one in this case.

While the gender-based clusterings that were so apparent in the MDS solution for voice similarity are not as obvious here in Figure 4.6, a clear separation of faces according to gender is still evident in the two-dimensional space. The diagonal line in Figure 4.6

highlights this pattern in the data; rotating the data so that this line is vertical clearly aligns the gender separation of faces with Dimension 1 on the horizontal axis, as I display on the following page in Figure 4.7.

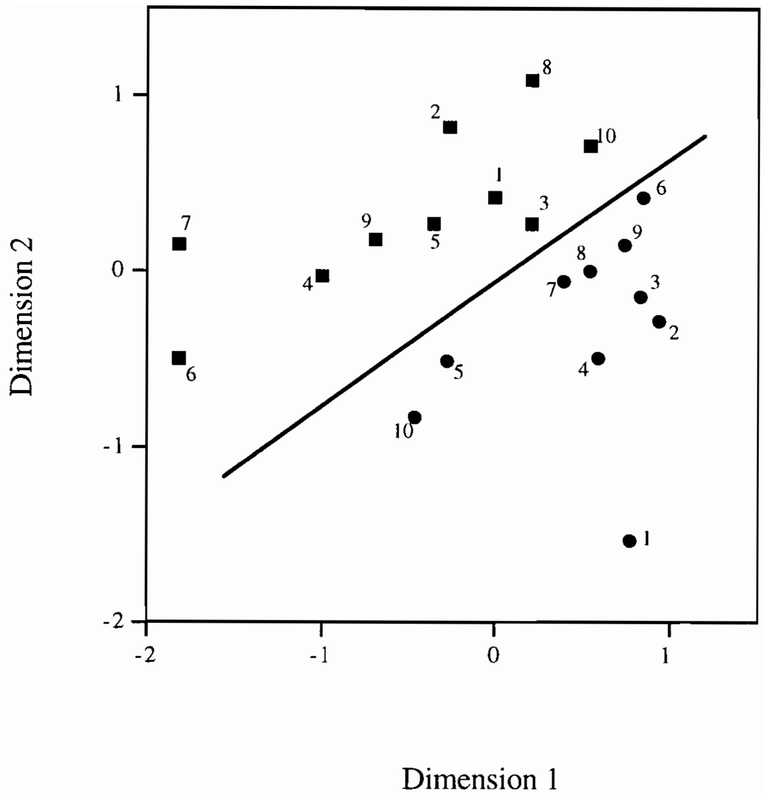

Figure 4.6: Two-dimensional scaling solution for face similarity of the face stimuli in Experiment 2. Male faces are indicated by filled squares, and female faces are indicated by filled circles. Numbers correspond to the specific face stimuli, which are displayed in Appendix C. We have added the diagonal line to indicate our subjective interpretation of Dimension 1, the general gender dimension, with all male faces located to the left of the line, and all female faces located to the right of the line.

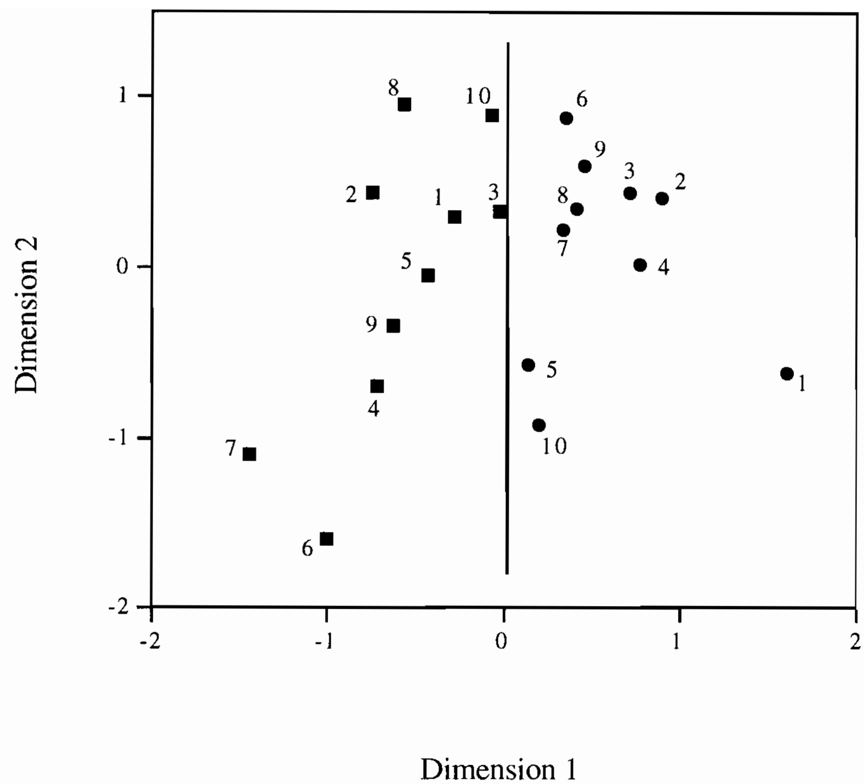

Figure 4.7: The rotated two-dimensional scaling solution for face similarity of the face stimuli in Experiment 2. Female faces are indicated by filled circles, and male faces are indicated by filled squares. Numbers correspond to the specific face stimuli, which are displayed in Appendix C. The vertical line indicates the division between male and female faces in Dimension 1, with male faces to the left of the line and female faces to the right of the line. The formulae for rotating the two-dimensional MDS space from its presentation in Figure 4.6 are modified from Edwards (1985:255-256), and are given in Appendix E.

While the similarity spaces for both male and female faces that are represented in Figure 4.6 and Figure 4.7 show much more diffusion than did the similarity space for voices, there still exists a clear-cut division between female and male faces. As was the case in the MDS solution for voice similarity, Dimension 1 for face similarity is also representative of gender. The more diffuse nature of the MDS results for face similarity, however, indicates that faces across gender categories are perceived as more similar to each other than are voices, and are therefore more confusable with each other. This was indicated by the participants' ubiquitous comment that the face similarity rating task was *much* more difficult for them than was the voice similarity task; stimuli of greater confusability lead to a more difficult perceptual task.

Taking note of the specific face stimuli that were employed in this similarity judgment task, as illustrated in Appendix C, it becomes apparent that the rotated Dimension 2 most likely represents skin hue, or some closely-related feature. The non-white faces in the set of 20 are grouped entirely near the bottom of the MDS spaces for both genders: of the male faces, Male Face 4, Male Face 6, Male Face 7, and Male Face 9 have skin of the darkest hue. Additionally, Male Face 1, Male Face 3, Male Face 8, and Male Face 10 have skin of the lightest hue (and likewise, appear at the top of the group). Of the female faces, Female Face 1, Female Face 5, and Female Face 10 have skin of the darkest hue, while Female Face 6 and Female Face 9 have skin of the lightest hue.

The Dimension 1 and Dimension 2 coordinates allow us to easily calculate which faces and voices exist at the centers of each of the MDS spaces; recall that the literature predicts that these exemplars will be the most stereotypical. In the next sections we assessed this notion, however, by collecting reaction time results for speeded voice and face gender classification tasks. As stereotypic voices and faces should be processed faster than those that are less stereotypic, we expected that our reaction time results would either confirm the aforementioned expectations about the exemplars within the MDS space by singling out the same voices and faces as the fastest to be processed, or shed doubt on them by singling out other voices and faces as fastest.

4.3 Experiment 3: Speeded Voice Gender Classification

4.3.1 Experiment 3 Purpose

Experiments 3 and 4 further assessed the results of the MDS scaling solutions that we arrived at in Experiments 1 and 2 above, examining whether there is a stereotype effect on the gender classification of voices and faces. Experiment 3 addresses whether the MDS topology found in Experiment 1 actually predicts gender classification, which is relevant for our selection of stereotypical voices and faces for the following experiments in Chapter 5. Audio tokens of voices uttering words were played to subjects, who were asked to classify the gender of the speaker as quickly as possible; that is, they were asked to answer the question "is this a man's or a woman's voice?" Reaction times (RTs) were compared across voices. Our main assumption here is that more-stereotypical voices will show shorter RTs for gender judgment, while less-stereotypical (i.e., nonstereotypical) voices will show longer RTs for gender judgment, due to increased processing load for nonstereotypical voices.

In information processing theory, increased processing load is reflected in longer RTs to stimuli, which are considered to show a "processing deficit." Since stereotypes are considered to serve as "cognitive shortcuts" in processing, any aspect of the input into cognition that varies from expectation should increase processing load. Therefore, nonstereotypical voices should be more complex for the cognitive system to process, resulting in increased RTs. The speeded identification tasks employed in Experiments 3 and 4 assessed this notion using the results of the MDS solutions described above.

4.3.2 Experiment 3 Methods

4.3.2.1 Participants

The participants in Experiment 3 were 10 Ohio State undergraduate and graduate students (2 males, 8 females; 5 undergraduates, 5 graduates). Some of the participants received partial course credit in their undergraduate linguistics courses in exchange for their participation, with the remaining participants receiving $5.00 for their participation. All were native speakers of American English, and all completed personal history questionnaires before taking part in the experimental session. None reported any speech or hearing disorders. All participants were debriefed immediately after their session.

4.3.2.2 Stimuli

The stimuli for Experiment 3 were the same 480 sound files of the 24 naturally-produced English words, each uttered by 20 speakers, that were used in Experiment 1 as described above in §4.1.3.2 on page 91. See Appendix A for a list of these words.

4.3.2.3 Procedure

Before starting Experiment 3, each participant read a printed list of experiment instructions in which they were told that they would be hearing a series of voices uttering different words, and that their task would be to judge the SEX of the speaker as quickly as possible. They were instructed to press the appropriately-labeled button on a response box for a *male* or *female* judgment, and were told that they did not need to wait until the end of the uttered word to enter their response.

Participants were tested in a sound-attenuated booth. MEL Professional software on a Dell 466/M computer was used to present the stimuli to participants and collect their responses. Participants heard the stimuli binaurally via Beyer Dynamic DT 220 High Fidelity headphones, and they logged their responses by pressing either the left-most or right-most button on a 5-button Psychology Software Tools serial response box. Of these two buttons, one was labeled *FEMALE* and the other was labeled *MALE*. For counter-balancing purposes, label position was alternated so that half of the subjects were tested with the left-most button labeled *FEMALE*, while the other half were tested with the right-most button labeled *FEMALE*.

Participants responded using their index fingers to push the buttons, and were instructed to keep their index fingers resting lightly on the two relevant response box buttons at all times throughout the experiment. Immediately after each response was logged, participants were presented with a visual display of their reaction time, in seconds. If participants failed to respond within 5 seconds, an indicator of "no response given" was logged. After each response was logged, an intertrial interval of 1000 msec occurred before the next voice stimulus was presented for gender judgment.

Participants first heard a practice block of 10 stimuli to familiarize them with the procedure for the speeded sex judgment task. All 480 of the sound file stimuli (24 words x 20 speakers) were then presented in random order to each participant, in 6 blocks of 80 stimuli each. Response latencies were measured from the start of the sound file (which corresponds with the onset of energy of the initial speech sound in the token word) until one of the response box buttons was pressed by the participant. Each experimental session lasted approximately 45 minutes.

4.3.3 Experiment 3 Results and Discussion

Average reaction time results for the speeded voice gender classification task are given for each speaker in Table 4.1. The percentage of times that each voice was accurately judged for gender is also given in Table 4.1. These results are averaged across all words and all participants for each speaker. (Reaction times that were further than 2 standard deviations from the mean RT for the entire session were identified as outliers and were excluded from the calculation of the results that are presented in Table 4.1)

Female Voices

Spkr No.	RT (msec)	%F
5	422.79	95.8
2	436.33	94.2
4	438.79	96.7
1	442.50	88.8
6	455.21	97.5
3	460.29	95.0
10	480.33	97.1
8	502.50	93.8
9	516.33	94.2
7	536.04	84.2

Avg. RT = 469.11

Male Voices

Spkr No.	RT (msec)	%M
5	461.25	95.8
1	475.79	95.8
7	476.67	95.8
3	477.88	97.1
2	480.17	96.7
4	488.50	97.9
10	491.75	96.3
6	493.08	98.3
9	499.63	95.4
8	578.50	96.7

Avg. RT = 492.32

Table 4.1: Reaction Time (RT) results for Experiment 3, *Speeded Voice Gender Classification*. Results for Female and Male speakers are separated into two lists, with RTs given in milliseconds in the second column of each list, and the list orders sorted from shortest to longest RT. Reaction time results are averaged across all participants' reactions to all 24 stimulus words for each speaker. The speaker identification number is shown in the first column of each list. The third column of each list gives the percentage of gender-accurate responses given for each speaker, calculated across all subjects' responses to all test tokens presented for that speaker. So, e.g., Female Speaker 5 was identified as "female" for 95.8 percent of the test tokens that presented her voice, and the average time that it took participants to make their gender judgment was 422.79 msec.

The reaction time results that are given in Table 4.1 indicate that Female Speaker 5 and Male Speaker 5 induced the shortest RTs for their respective gender groups, while Female Speaker 7 and Male Speaker 8 induced the longest RTs. It is interesting to note, however, that neither Female Speaker 5 nor Male Speaker 5 showed the highest percentages of gender-accurate responses for their respective groups. But Female Speaker 7, with the longest RT for the female speakers, did also show the lowest percentage of gender-accurate responses for the female voices, however. This indicates that Female Speaker 7 was the most difficult to categorize with respect to gender, and the slower average RT here is associated with greater confusability for this voice.

There is an interesting asymmetry between the female and male voices when it comes to accuracy of gender judgment. While all of the male voices show gender judgment accuracy scores in the mid- to high-90 percent range, there is much more apparent variation in the percent-correct accuracy judgments for the female voices.

Figure 4.8 displays voice gender identification RTs for each speaker graphed against the average word duration lengths for each speaker. Data values for this display are given in the following Table 4.2. As there is some apparent variation in average word duration across speakers, which indicates slightly varying speaking rates, we must ask whether this variation should be taken into account when considering the RTs evoked by each speaker in the speeded gender classification task. The results of a regression on gender identification RTs against average word duration indicate that a large amount of the variability in the gender identification RTs is predicted by word duration variability (R^2 = .994; DF = 1, 19; $p < .01$). Therefore, we will need to keep the effect of duration in mind when evaluating gender identification RTs as a measure of voice stereotypicality.

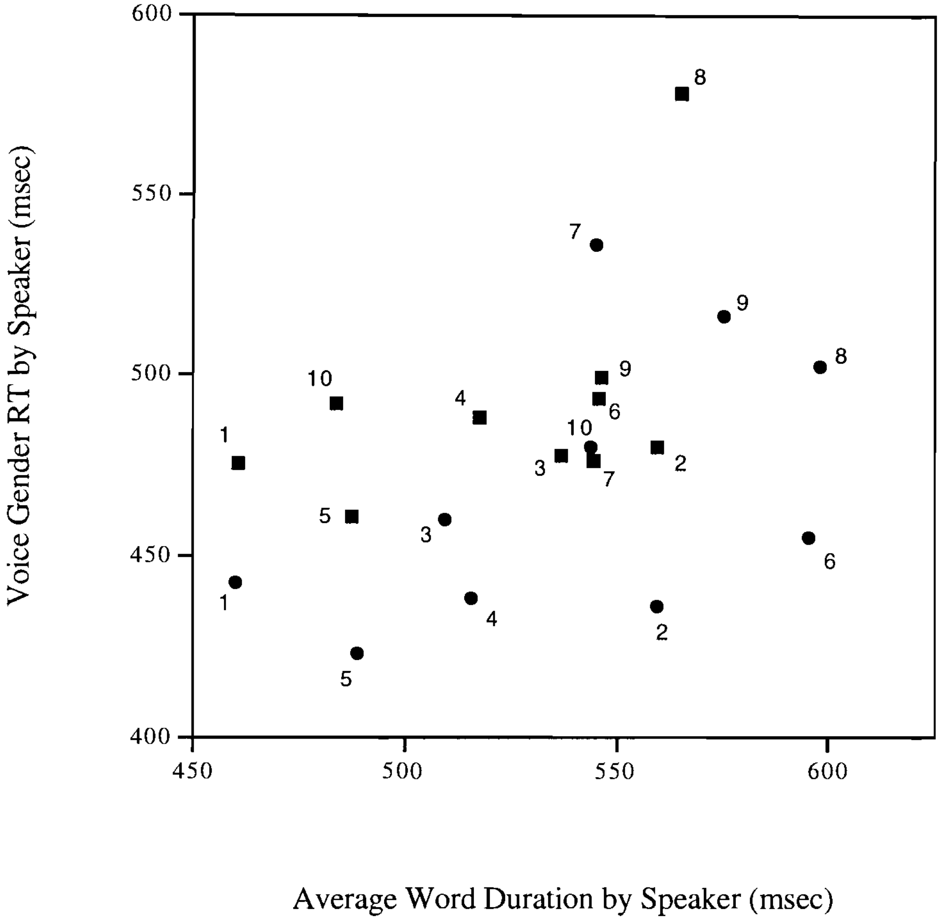

Figure 4.8: Voice gender RT by average word duration for each speaker, with time given in milliseconds. Female speakers are indicated by filled circles, and male speakers are indicated by filled squares. The numbers that mark each token correspond to the specific speaker for which the RT was calculated. Values for each data point are given in Table 4.2 on the following page.

Female Voices

Spkr No.	RT (msec)	Word Dur.
1	442.50	460.00
2	436.33	559.79
3	460.29	509.38
4	438.79	515.50
5	422.79	489.04
6	455.21	595.54
7	536.04	545.46
8	502.50	597.79
9	516.33	575.71
10	480.33	543.75

Male Voices

Spkr No.	RT (msec)	Word Dur.
1	475.79	460.25
2	480.17	559.63
3	477.88	537.25
4	488.50	517.67
5	461.25	487.42
6	493.08	545.92
7	476.67	544.75
8	578.50	565.13
9	499.63	546.63
10	491.75	483.79

Table 4.2: Voice gender RT values and average word duration values for the data points displayed in Figure 4.8. All values are given in milliseconds, and the values given for voice gender RT are repeated from Table 4.1. Word duration values were computed by averaging the duration of each of the 24 words as spoken by each of the 20 speakers.

4.4 EXPERIMENT 4: SPEEDED FACE GENDER CLASSIFICATION

4.4.1 Experiment 4 Purpose

In Experiment 4, we further assessed the results of the MDS scaling solution for faces from Experiment 2 by recording reaction times in a speeded gender classification task for faces. Again, we expect that the increased processing load associated with nonstereotypic face stimuli would be reflected in longer RTs to those stimuli, versus shorter RTs to stereotypic face stimuli.

4.4.2 Experiment 4 Methods

4.4.2.1 Participants

The participants in Experiment 4 were 10 Ohio State undergraduate and graduate students (8 females, 8 males; 4 undergraduates, 12 graduates). Some of the participants received partial course credit in their undergraduate linguistics courses in exchange for their participation, with the remaining participants receiving $5.00 for their participation. All were native speakers of American English, and all completed personal history questionnaires before taking part in the experimental session. None reported any speech or hearing disorders. All participants were debriefed immediately after their session.

4.4.2.2 Stimuli

The stimuli for Experiment 4 were the same 20 picture files of the 10 female and 10 male faces that were used in Experiment 2, as described above in §4.2.3.2 on page 102. See Appendix C for a display of the face stimuli.

4.4.2.3 Procedure

Before starting Experiment 4, each participant read a printed list of experiment instructions in which they were told that they would be seeing a series of people's faces, and that their task would be to judge the SEX of the person they saw as quickly as possible. They were instructed to press the appropriately-labeled button on a response box for a *male* or *female* judgment.

Participants were tested in a sound-attenuated booth. MEL Professional software on a Dell 466/M computer was used to present the stimuli to participants and collect their responses. The visual stimuli were shown to participants on a computer monitor, and were of the same display size as they were in Experiment 2, as well as again appearing on a black background. Participants logged their responses by pressing either the left-most or right-most button on a 5-button Psychology Software Tools serial response box. Of these two buttons, one was labeled *FEMALE* and the other was labeled *MALE*. For counterbalancing purposes, label position was alternated so that half of the subjects were tested with the left-most button labeled *FEMALE*, while the other half were tested with the right-most button labeled *FEMALE*.

Participants responded using their index fingers to push the buttons, and were instructed to keep their index fingers resting lightly on the 2 response box buttons at all times throughout the experiment. Immediately after each response was logged, participants were presented with a visual display of their reaction time, in seconds. If participants failed to respond within 5 seconds, an indicator of "no response given" was logged. After each response was logged, an intertrial interval of 1000 msec occurred before the next face stimulus was presented for gender judgment.

Six blocks of the 20 faces in random order were presented to each participant for sex judgment. Response latencies were measured from the initial presentation of each face picture until one of the response box buttons was pressed by the participant. The sessions for Experiment 4 lasted approximately 15 minutes, prior to which each participant had also taken part in Experiment 5 (as described in Chapter 4), for a total session time of approximately 30 minutes.

4.4.3 Experiment 4 Results and Discussion

Reaction time results for the speeded face gender classification task are given for each speaker in Table 4.3. These results are averaged across the responses of all participants.

FEMALE FACES

Face No.	RT (msec)
3	440.44
6	444.33
8	445.52
5	449.16
10	449.64
1	451.58
7	452.05
4	454.26
2	462.14
9	462.89

Avg. RT = 451.20

MALE FACES

Face No.	RT (msec)
7	440.04
6	448.86
9	455.27
2	458.84
4	461.39
1	483.15
8	486.98
10	487.60
5	493.17
3	554.24

Avg. RT = 476.95

Table 4.3: Reaction Time (RT) results for Experiment 4, *Speeded Face Gender Classification*. Results for Female and Male speakers are separated into two lists, with RTs given in milliseconds in the second column of each list. RT results are averaged across all participants' reactions to all 20 faces. The face identification number is shown in the first column of each list. Each face was accurately classified for gender 100 percent of the times that it was presented.

The results presented in Table 4.3 indicate that Female Face 3 and Male Face 7 induced the shortest RTs for their respective gender groups, while Female Face 9 and Male Face 3 induced the longest RTs. Also, it is evident that there is a much greater amount of variability in reaction times to the male faces (with a spread of more than 100 msec between most quickly reacted-to and most slowly reacted-to faces), versus the female faces (with only a 20 msec spread between fastest and slowest).

Note that Male Face 7 and Male Face 6, with the two shortest reaction times, were judged as most dissimilar to the female faces along Dimension 1 in Figure 4.7 on page 108. Also, Male Face 2 was the furthest to the left of all the other female faces in Figure 4.7, other than Male Faces 6 and 7.

It is also worth noting that gender judgment accuracy was at ceiling for all of the faces; recalling that there was a good amount of variation in gender judgment accuracy for the voices, especially among the female voices, it is clear that gender was more accurately perceived for faces than for voices. This is especially interesting in light of the more diffuse MDS results for face similarity judgments—while faces were more confusable in the similarity judgment task, they were more easily identified as male or female in the speeded gender classification task. Taken together, though, this may simply be an indication that face stereotypicality varied less than voice stereotypicality for the stimuli that we employed.

4.5 SELECTING VOICES AND FACES FOR FURTHER STIMULI CREATION

The results from Experiments 1, 2, 3, and 4 provide us with three types of information to use in the selection of stereotypical and nonstereotypical voices and faces for creation of the stimuli to be used in the speech perception experiments that I describe in Chapter 5. First, Experiments 1 and 2 provide us with multidimensional scaling solutions for judged similarity between faces and voices. Second, Experiments 3 and 4 provide us with reaction time results for speeded gender classification of faces and voices. And finally, we can also consider accuracy for the gender classification tasks, which is useful for interpreting our results in general.

Prior research suggests that data points near the centers of clusters in MDS solutions are the most typical, which in terms of cognition are considered the most stereotypical (see Busey 1998). We examined this notion by including the RT results of the speeded gender classification tasks, on the assumption that faster RTs (and therefore faster processing) are indicative of more stereotype-like stimuli. So, the MDS and RT results were used in conjunction to select the voices and faces meeting our stereotypicality criteria for use in stimuli construction in the Chapter 5 experiments.

4.5.1 Selecting Voices

Here, to select the voices that we would consider to be "stereotypical," we considered the MDS results from Experiment 1 in conjunction with the RT results from Experiment 3, selecting and voices that were near the centers of their respective MDS gender groupings, but which also had short RTs for their gender categories. (In fact, in the end we relied almost exclusively on shortest RTs to select our voices, based on the notion that stereotypic exemplars will be processed more quickly. In this way, we could compare our RT results with the expectation based on the aforementioned notion that exemplars near the centers of clusters in an MDS space will be the most prototypical.)

Based on these criteria, the female voice that we chose as stereotypical was Female Speaker 5, and the male voice that we chose as stereotypical was Male Speaker 5. The reaction time results, as given in Table 4.1, generally confirm our expectations about the central location of these typical exemplars in the MDS groups (i.e., both Female Speaker 5 and Male Speaker 5 are near the centers and induce the shortest RTs), even with the relatively small number of voices that were involved in the similarity judgment task.

The discussion in the literature regarding the ways in which non-typical exemplars might pattern within MDS spaces has been little more explicit than simply noting that these non-typical members will be found "away" from the centers of groups. Before conducting Experiments 1 and 3, we did not know what this would imply for the locations in which we would discover our most nonstereotypical voices: would the nonstereotypical male voice, for example, be the exemplar that was located furthest from the cluster of female voices (i.e., on the "other side" of the stereotypical male exemplar), or would the nonstereotypical male voice exist nearest the female voice cluster (reflecting, perhaps, more "confusability" with the female voices)?

In the case of selecting the nonstereotypical voices, we relied mostly on the RT results from the gender classification task in Experiment 3 to inform our selection of the two voices. The longest gender classification RTs were induced by Female Speaker 7 and Male Speaker 8, as shown in Table 4.3. When these results are compared with those for the voice similarity MDS solution shown in Figure 4.4, it is interesting to note that both Female Speaker 7 and Male Speaker 8 appear at the bottoms of their respective gender groups, as relative outliers. So, Dimension 2 in the MDS mapping for voice similarity seems more important for stereotypicality even though Dimension 1 separated the gender groups better.

Again, the MDS and RT results led us to the same conclusion in the selection of these two voices as nonstereotypical. As I mentioned briefly in an earlier section of this chapter, it is likely that Dimension 2 of this MDS solution represents some aspect of femininity/masculinity, or perceived "gender-appropriateness" in terms of perceived voice quality, with the two voices that are judged as furthest from the gender ideal for this quality existing at this extreme of the dimension. (The relatively low percentage of gender-accurate judgment for Female Speaker 7 (84.2%) lends support to the notion that this voice was less confidently judged as female. Note that Female Speaker 1, located in close proximity to Female Speaker 7 in the MDS mapping, also shows a relatively low gender accuracy percentage, 88.8%).

It should be noted, however, that it is not necessary for Dimension 2 to be representing exactly the same quality in both the female and male voices; it could be the case that the second-most salient quality for judging female voices is actually different from that employed in judging male voices, which might account for the variation in diffuseness between the male and female voices that is evident in the MDS mapping. Murry and Singh (1980), for example, suggested that listeners used different dimensions depending on the gender of the speaker, and interpreted their MDS results as indicating that listeners judged female voices primarily along a breathiness continuum, but did not use a similar dimension in judging male voices.

Figure 4.9 graphically presents the relationship between gender identification RT and distance from the center of the gender clusters for our voice similarity ratings. In order to better visualize the relationship between our gender identification RT results (which we are assuming to indicate stereotypicality) and the centers of the clusters in MDS space (assumed in the literature to indicate prototypicality), we calculated the exact "center" locations for each gender cluster in the two-dimensional voice similarity space by first finding the median values in both dimensions for each cluster. We then calculated the Euclidean distance of each voice exemplar from the exact center of its respective cluster.

Figure 4.9 clearly indicates the trend for voice exemplars that are further from the centers of their clusters to be reacted to more slowly. And indeed, the relationship between RT and Euclidean distance from the center of the space is significant; in a regression of RT against the Euclidean distance from the center of the gender cluster, the distance predicted nearly 70 percent of the variability in RT ($R^2 = .678$; DF = 1, 19; $p < .001$). As I mentioned earlier, it is also necessary to consider the influence of speech rate on these results, as well; by including word duration as a second independent variable in a multiple regression, results indicate that these two variable account for almost all of the variability in RT ($R^2 = .995$; DF = 2, 18; $p < .001$).

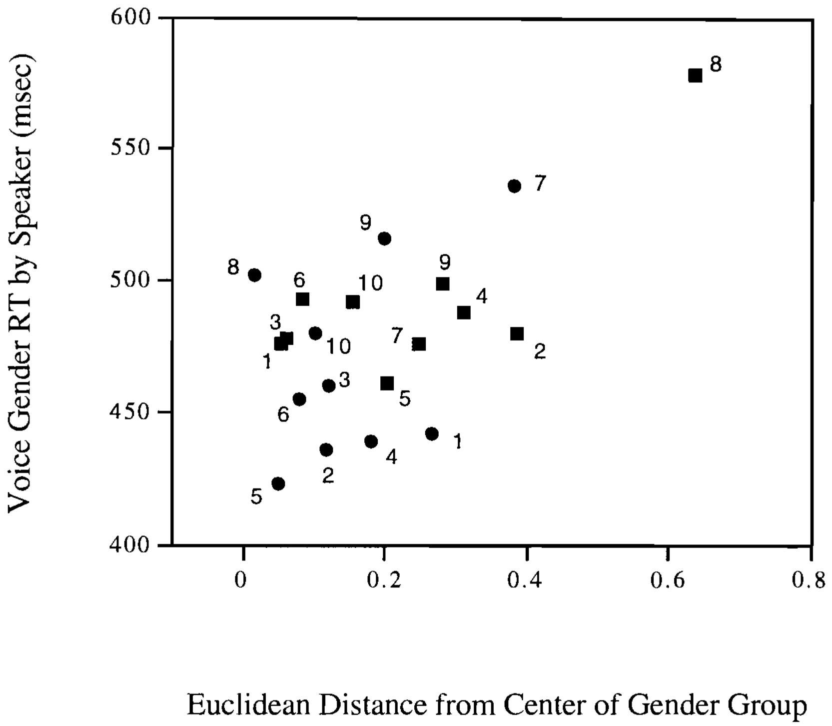

Figure 4.9: Voice gender RT for each speaker plotted against the speaker's Euclidean distance from the center of the voice gender cluster in the MDS map space. RT is represented in msec along the vertical axis, and the horizontal axis represents distance from the center of each gender cluster in the MDS voice space; distance units are arbitrary, and are relevant only to the MDS space, with "0" indicating the median point. Female speakers are indicated by filled circles, and male speakers are indicated by filled squares. The numbers that mark each token correspond to the specific speaker for which the RT was calculated. RT values for each data point are given in Table 4.2.

4.5.2 Selecting Faces

Selecting the stereotypical and nonstereotypical faces for use in further stimuli construction was a bit less clear-cut, in the sense that we were unable to rely on the distinct gender groupings that had resulted from the MDS mapping for voices; in the case of the face similarity judgment task in Experiment 2, as is graphically displayed in Figure 4.6 and Figure 4.7, there was a much smaller separation between male and female faces in the similarity mapping. While there is still a definite category boundary between genders (as we have represented by inserting the Dimension 1 division line in each graphic), the variability within each gender group is much greater than the distance between the median values for the two groups. This makes it difficult to rely on our original heuristic of selecting an exemplar that is located at the center of a distinct group as most stereotypical.

So in the case of selecting stereotypical and nonstereotypical faces, we relied primarily on the speeded gender classification RT results from Experiment 4, which are given in Table 4.3. Based on these results, we selected the female face with the shortest RT, Female Face 3, as our stereotypical face, and the female face with the longest RT, Female Face 9, as our nonstereotypical face.

When we examined the RTs to the male faces, we found much more varability in RTs than was evident for the female faces. We also encountered a slightly confounding situation when we examined the faces with the shortest reaction times in our attempt to select the stereotypical male face: the three male faces that had induced the shortest reaction times, Male Face 7, Male Face 6, and Male Face 9, were the darkest faces (see Appendix C). This suggested to us that the reaction times for these three faces could have at least partially resulted from a response strategy employed by the participants, i.e., "see a non-white face, respond MALE right away." To avoid confounding our further results, we elected to eliminate these three faces at this time, and selected Male Face 2 (with the next-shortest RT) to represent the stereotypical male face.[5]

We selected the male face with the slowest RT, Male Face 3, as our nonstereotypical male face. When we compared our four face selections, based on the RT results from Experiment 4, with the results of the MDS solution for faces from Experiment 2, a nice pattern among the four faces emerged in Figure 4.7: the two faces that we had selected as our stereotypical female and male exemplars (Female Face 3 and Male Face 2) are further from the Dimension 1 gender boundary between faces (and near the median values for their groups), and the nonstereotypical exemplars (Female Face 9 and Male Face 3) are closer to the boundary. For faces, it seems that the most stereotypical male and female exemplars exist further from each other, and are therefore less confusable, with nonstereotypical exemplars existing within the more confusable intermediate similarity space.

It is interesting to note that while we found what we consider to be a response strategy effect for non-white male faces, there was no such effect for female faces—in this case, the non-white Female Face 1, Female Face 5, and Female Face 10 induced reaction times that were at the middle of the RTs for the entire female group. Perhaps different strategies are used in processing female faces versus male faces, even though both genders are perceived with 100% accuracy.

---

[5] It will, however, be very important for us to conduct further analysis on this point, because it is reasonable to consider that perceived race/ethnicity via faces might be a meaningful component of the stereotypes that our participants hold. It is interesting to note here that we found faster RTs for the darker male faces, but NOT for the darker female faces (who actually induced *slower* RTs). This could indicate that for our collection of MALE faces, the African American and Middle Eastern American males appear *more* stereotypically male (i.e., darker faces make "maler males"), while for our collection of FEMALE faces, the non-white females are perceived as *less* stereotypically female (i.e., lighter faces male "more feminine females"). Therefore, it is quite possible that ethnicity may be interpreted differently as a component of gender, across genders, at least for American culture. This is a point that warrants further examination.

Figure 4.10 presents a graphic display of the (rotated) Dimension 1 values for the MDS face similarity space from Experiment 2, plotted again the reaction times for face gender classification from Experiment 4. While reactions times for female *voices* showed much more variability than did the reactions times for male *voices,* in the case of faces this trend is reversed: there is a much greater span apparent in the reactions times for the male faces than for the female faces. Additionally, the male faces that are located closer to the center of the perceptual space in the rotated Dimension 1 (as indicated by point "0" on the horizontal axis), show increasingly longer reaction times, perhaps indicating that the male faces closer to the center of the space are more confusable, and generally less prototypical.

But note that the same is not true for the female faces. This could mean simply that race/ethnicity is more of a component of male stereotypicality, or that these male faces (as compared to the male voices) show a greater range of stereotypicality values.

While Figure 4.10 presents RTs according to distance from the center of the collective cluster of faces in terms of Dimension 1, Figure 4.11 provides a graphic display of face gender identification reaction times plotted against the Euclidean distance of each face exemplar from the center of its respective gender group. Euclidean distance values were calculated in the same fashion as they were for the voices in §4.5.1. A regression of face gender RT against Euclidean distance from the center of the gender cluster indicated that this relationship is significant ($R^2 = .66$, DF = 1, 19, $p < .001$). Therefore, faces that lie further from the centers of their respective gender clusters are slower to be processed. These RT results support the indication in the literature that stereotypic exemplars lie at the centers of MDS clusters.

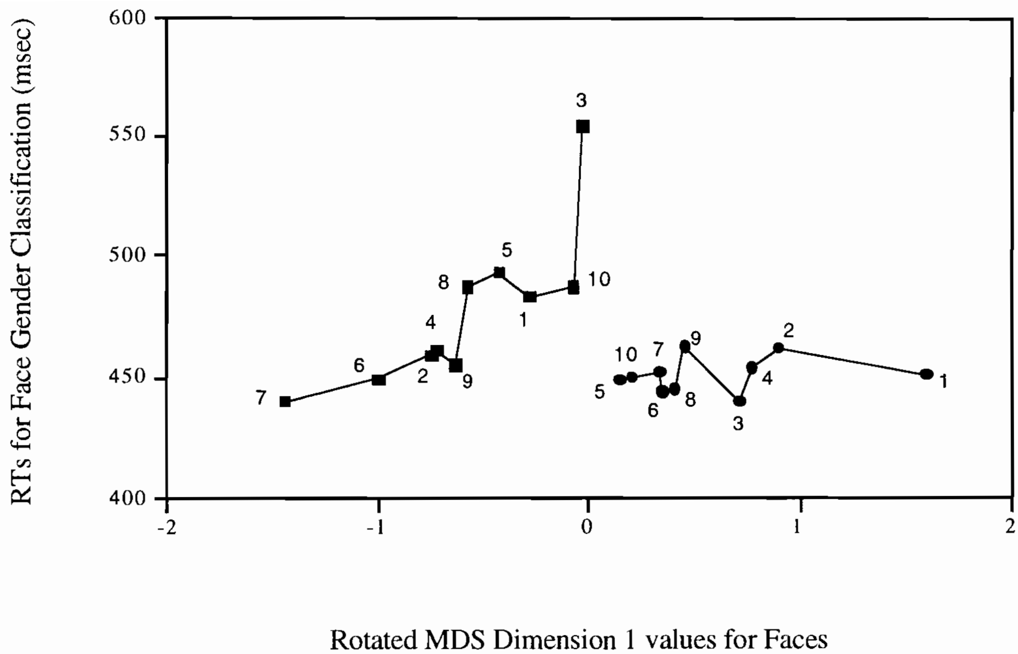

Figure 4.10: Display of the rotated MDS Dimension 1 values for face similarity from Experiment 2 (on the horizontal axis), plotted by the RTs in milliseconds for speeded face gender classification from Experiment 4 (on the vertical axis). Female faces are indicated by filled circles, and male faces are indicated by filled squares.

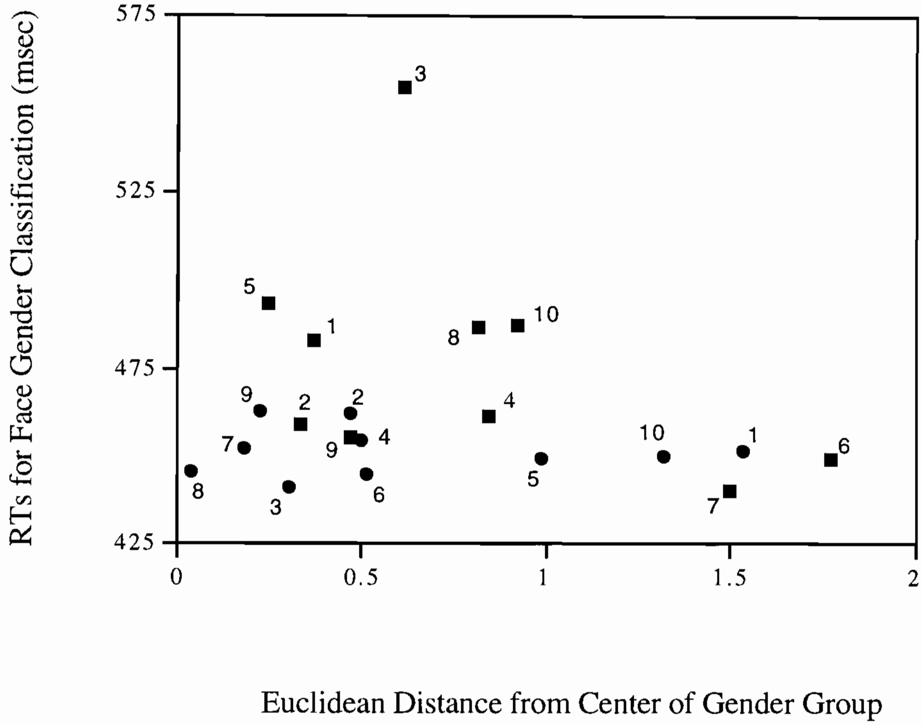

Figure 4.11: Face gender RT for each face plotted against the exemplar's Euclidean distance from the center of its respective face gender cluster in the MDS map space. Female faces are indicated by filled circles, and male faces are indicated by filled squares. The numbers that mark each token correspond to the specific face for which the RT was calculated. RT values for each data point are given in Table 4.3, and the face gender exemplars are displayed in Appendix C.

4.6 SUMMARY

In this chapter, I have discussed several methods that we employed in our quest to tap implicit assessments about stereotypic and nonstereotypic voices and faces of speakers. The confluence of three types of results provides us with a novel means for determining stereotypicality of voices and faces. These three types of results include: (1) multidimensional scaling solutions for voice and face similarity, (2) reaction times in speeded gender classification tasks for faces and voices, and (3) percent-accuracy of gender judgments in the same gender classification tasks.

We identified two voices and faces as stereotypical, and two voices and faces as nonstereotypical, for use in the stimuli construction that I describe in Chapter 5. These faces and voices are as follows in Table 4.4:

	FEMALES	MALES
STEREOTYPICAL VOICES	Female Speaker 5	Male Speaker 5
NONSTEREOTYPICAL VOICES	Female Speaker 7	Male Speaker 8
STEREOTYPICAL FACES	Female Face 3	Male Face 2
NONSTEREOTYPICAL FACES	Female Face 9	Male Face 3

Table 4.4: Stereotypical and Nonstereotypical faces and voices selected for stimuli creation in Chapter 5, based on the results of the gender perception experiments in this chapter. See Appendix C for representations of the female and male faces.

CHAPTER 5

GENDER STEREOTYPES AFFECT SPEECH PROCESSING: AUDITORY NAMING AND VISUAL PRIMING RESULTS

5.0 INTRODUCTION

Having chosen stereotypical and nonstereotypical faces and voices on the basis of the implicit measures that I discussed in the previous chapter, we are now prepared to address the question of whether stereotypes about gender influence low-level speech perception. The two experiments in this chapter shed light on this question.

While earlier social psychological conceptions of the cognitive nature and function of stereotypes assumed that stereotypes were mostly consciously controlled in processing, the tides have turned; much of the work from the past two decades has indicated that indeed many if not most components of stereotypic responses occur outside of consious control. As Banaji and Hardin (1996:136) pointed out, "stereotyping can occur implicity, without subjects' conscious awareness of the source or use of stereotypic information in judgment."

In order to assess the effects of implicit stereotyping on further processing, social psychologists have adapted classic priming procedures to test whether processing can be speeded through the use of certain types of stereotypic primes (see §5.2 below for a discussion of the priming procedure in general). As Fiske (1998:368) stated:

> People [process] stereotype-consistent information faster, presumably because it is easier to assimilate . . . *People encode less perceptual information when they can use stereotypes and expectancies to fill in the details.* [emphasis added]

This quotation from Fiske captures the essence of symbolic processing. It also implies that the influence of stereotypic primes should be evident in decreased reaction times to associated targets, as the primes will activate stereotypes that will then ease the target's processing burden (again, see §5.2 below).

A number of researchers have begun to examine the ways in which gender stereotype effects occur automatically and implicitly to affect processing, which is directly relevant to the effects that we are interested in examining in this chapter. Banaji and Hardin (1996), e.g., presented subjects with either stereotypically-gendered word primes salient in American culture (e.g., "nurse," "mechanic"), or neutral word primes (e.g., "student," "person"). Primes were followed by pronoun targets that were either gender-related (e.g., "she," "he") or neutral (e.g., "it," "me"). Subjects' task was to determine whether the target was male or female. The subjects were also assessed on their explicit beliefs regarding gender stereotypes, the reform of gender-exclusionary language, and the influence of gender in peoples' lives. The researchers concluded that "automatic gender beliefs (stereotypes) were observed in faster responses to pronouns consistent than inconsistent with the gender component of the prime regardless of subjects' awareness of the prime-target relation, and independently of subjects' explicit beliefs about gender stereotypes in language reform" (p. 136). So in Banaji and Hardin's experiments, gender stereotypes did implicitly influence processing.

We adapted this priming technique for our own exploration of the effects of gender stereotypes that are evoked by faces on speech perception. The voices and faces selected through the experiments in Chapter 4 were first used to create audio-only and multimodal tokens for use in the speech perception experiments in this chapter, which employ matched and mismatched mixtures of faces and voices in terms of stereotypicality. The first experiment described here involves an auditory naming task in which we employed audio test tokens to assess whether words produced by gender-stereotypic voices are easier to process. In this naming task, listeners were presented with speech tokens and asked to repeat the word that they heard as quickly as they could. The question that we are asking is whether it takes listeners less time to process and repeat the word they hear when the voice producing the token is more stereotypical.

The second experiment involves an auditory naming task with audiovisual test tokens to assess whether there is an interaction of face and voice stereotypicality in speech processing. Again, listeners were presented with speech tokens for which the speakers varied in gender stereotypicality, but this time the listeners also were primed with a picture of the face of the presumed speaker. Our prediction was that we would find an interaction between voice and face stereotypicality, so that audiovisual tokens with faces and voices that are matched for stereotypicality will be processed faster than those that are unmatched for stereotypicality, and stereotypical voices in general will be processed faster than nonstereotypical voices.

In this chapter, I first I discuss our audio-only speech perception Experiment 5 in §5.1. In §5.2, I provide more details on the standard priming procedure. And finally, in §5.3, I discuss Experiment 6, where we examined speech perception when it is primed with pictures of the faces of speakers. These two experiments provide support for the claim that gender stereotypes do indeed affect low-level speech perception processing.

5.1 EXPERIMENT 5: SPEEDED NAMING TASK (AUDIO-ONLY)

5.1.1 Experiment 5 Purpose

In this experiment, participants were presented with sequences of audio-only tokens of the four different voices chosen on the basis of the responses to Experiments 1 and 3, saying the words given in Appendix A, and were told to repeat the words as quickly as possible. This experiment provided us with more evidence regarding the question of whether speech produced by nonstereotypic voices is more difficult to process. Recalling that stereotypes serve as cognitive shortcuts in processing incoming stimuli, our prediction was that the more stereotypic the voice, the faster the response time would be. (Accuracy was not be examined in this experiment, only RT, although accuracy data could be interesting to examine in the future.)

This experiment is different from the speeded gender identification tasks described in Chapter 3 in that here, the input to the perceptual mechanism must be linguistically decoded, so that the focus is taken off of the specific speaker and placed onto the words that must be perceived. We hypothesized that stereotypes that are activated by characteristics of the voice will interact with the mechanism perceiving the words to influence reaction time. The methods employed in this experiment were based on those described in Mullennix, Pisoni, and Martin (1989), who used an auditory naming task (also known as "speeded shadowing") to examine the effects of talker variability on spoken word recognition.

5.1.2 Experiment 5 Methods

5.1.2.1 Participants

The participants for Experiment 5 were 24 Ohio State undergraduate and graduate students (14 females, 10 males; 8 undergraduates, 16 graduates). Some of the participants received partial course credit in their undergraduate linguistics courses in exchange for their participation, with the remaining participants receiving $5.00 for their participation. All were native speakers of American English, and all completed personal history questionnaires before taking part in the experimental session. None reported any speech or hearing disorders. All participants were debriefed immediately after their session. (Ten of the participants for Experiment 5 also served as participants in Experiment 4: Speeded Gender Classification of Faces, as described in Chapter 3).

5.1.2.2 Stimuli

The stimuli for Experiment 5 were 192 digitized sound files, comprising 48 words as uttered by each of the 4 speakers who were selected in §4.6 of the previous chapter (Female Speaker 5, Female Speaker 7, Male Speaker 5, and Male Speaker 8). The sound files for each of these speakers' utterances of the 24 words that were originally used as stimuli in Experiments 1 and 3 of the previous chapter were used again as stimuli here. Additionally, 24 new words that were selected in the same manner as the first 24 were digitized and included as stimuli as well, following the selection and digitization procedures described in §4.1.3.2 on pages 91-94. The words that were selected for inclusion in these experiments were controlled for lexical density, initial phonetic segment, and neighborhood frequency, so that potential lexical and segmental effects could be minimized (see §4.1.3.2 for more information). These additional 24 words had been included on the word lists that each speaker read at his or her original audio—video tape-recording session, so that all natural utterances for construction of the stimuli used in all the experiments described in Chapters 4 and 5 were recorded at the same time.

See Appendix F for a list of these 48 words.

5.1.2.3 Procedure

Before starting Experiment 5, each participant read a printed list of experiment instructions in which they were told that they would be hearing a series of voices uttering different words, and that their task would be to repeat back the word that they heard as quickly and accurately as possible.

Participants were tested in a sound-attenuated booth. They were first fitted with a Sennheiser HMD 410 Headset/Boom Microphone Set, which is an integrated headphone and microphone set that allows for both playback of sound to the participants as well as the recording of their speech. The microphone boom was adjusted so that the head of the microphone was positioned toward the left edge of the participant's mouth, approximately one-half inch away from the lips.

MEL Professional software on a Dell 466/M computer was used to present the stimuli to participants and collect their responses. The participants heard each stimulus binaurally via the Sennheiser HMD 410 set's headphones, and repeated the target word back into the set's microphone as soon as they could identify it. The participants' input into the microphone was first fed into a TTE 411AFS Amplifier for preamplification, set for a 10K lowpass filter. The output from the amplifier was then fed into the microphone input jack of a 5-button Psychology Software Tools serial response box. Speech picked up by the microphone activated the box's sound-triggered switch and caused the participant's response time to be logged in a computer data file.

Immediately after each response was logged, participants were presented with a visual display of their reaction time, in seconds. If participants failed to respond within 4 seconds, an indicator of "no response given" was logged. After a response was logged, an intertrial interval of 1000 msec occurred before the next stimuli set was presented.

Participants first heard a practice block of 20 stimuli to familiarize them with the procedure for the speeded naming task. All 192 of the stimulus sound files (48 words x 4 speakers) were then presented in random order to each participant, in 4 blocks of 48 stimuli each. Each speaker was heard 12 times in each block, and each of the 48 words was heard once in each block. Response latencies were measured from the start of the sound file (which corresponds with the onset of energy of the initial speech sound in the token word) until the participant's vocal response triggered the timing circuit in the response box. Each experimental session lasted approximately 20 minutes.

5.1.3 Experiment 5 Results and Discussion

The reaction time results for the Speeded Naming Task (Audio-Only) for each speaker, averaged across the responses of all participants, are as follows: Stereotypical Male Speaker = 631.26 msec; Stereotypical Female Speaker = 631.30 msec; Nonstereotypical Male Speaker = 684.79 msec; and Nonstereotypical Female Speaker = 653.34 msec. These results are graphically presented in Figure 5.1.

Using Systat, version 5.2.1, software, we conducted a repeated-measures analysis of variance (ANOVA) on the reaction time data, with the two factors "voice gender" (female speaker or male speaker) and "voice stereotypicality" (stereotypical speaker or nonstereotypical speaker). Two significant main effects were obtained as the results of the ANOVA. First, there was a significant main effect of voice gender [$F\ (1,23) = 13.638$; $p<.001$], with the female voices as a group inducing shorter reaction times. The display in Figure 5.1 indicates that a good portion of this effect, however, can be explained by the Nonstereotypical Male Speaker's much longer average RT result. I will return to this issue below.

The second significant effect that we identified is the main effect of voice stereotypicality [$F\ (1,23) = 183.866$; $p<.001$]. This means that the stereotypical voices were processed and reacted to significantly faster than the nonstereotypical voices, and the large F value indicates that this effect is very robust. Indeed, the effect of voice stereotypicality is much more robust than the effect of voice gender.

We also found a significant interaction between voice gender and voice stereotypicality [$F\ (1,23) = 75.699$; $p<.001$], which is evident in the large difference in the slope of the lines connecting the data points that are displayed in Figure 5.1. Recall that the Nonstereotypical Male Speaker (AKA Male Speaker 8) showed a much longer average word duration than did the Stereotypical Male Speaker (AKA Male Speaker 5; duration values of 565.13 msec versus 487.42 msec, respectively, as displayed in Table 4.2 of Chapter 4). The Stereotypical Male Speaker was in fact the male voice with the longest average word durations, as well as the longest RT in the gender classification task. The Nonstereotypical Female Speaker (AKA Female Speaker 7) also had longer word durations, but she was not the slowest female speaker (although her RTs were longest). It is possible that the longer word durations for the two nonstereotypical voices (reflecting a slower speech rate) were contributing significantly to the main effect of stereotypicality, and complicating our interpretation of the gender stereotypicality interaction. This was something that was necessary for us to try to tease apart.

To test for the potential influence of average speaker word duration, we conducted a follow-up analysis of covariance (ANCOVA), which is a statistical analysis consisting of the combined application of linear regression and analysis of variance techniques (Freund & Williams 1966:3). Analysis of covariance can be used when treatments are compared in the presence of concomitant variables which can be neither eliminated nor controlled, which in this case is the variable (or covariate) of word duration. (Again, we used Systat, version 5.2.1, software, to compute the ANCOVA.)

The results of our ANCOVA application indicated that word duration did indeed account for a significant amount of the variance in the RTs to each of the voices [$F\ (1,187) = 42.938$; $p<.001$]. We also again found a significant main effect of voice stereotypicality [$F\ (1,187) = 4.705$; $p=.031$]; while this effect appears to be less robust when word duration is considered as a covariate, it is still significant. The results of the ANCOVA, however, showed neither the significant main effect of voice gender nor the interaction of stereotypicality with voice gender that were present in our earlier by-subjects

ANOVA results. This indicates that the voice stereotypicality effect is real and robust, and is independent of any word duration effect, but the voice gender effect and gender–stereotypicality interactions are entirely due to the longer RTs induced by the nonstereotypical male voice. While the nonstereotypical female voice was slow, the nonstereotypical male was at the extreme for rate for his gender group, suggesting that male voices are stereotypically faster—perhaps the same reason that female voices in American culture are stereotypically more hyperarticulate. Also, this may indicate that female speakers have a greater range of voice styles (i.e., a greater range of voice-gender roles) available to them.

So, the results of Experiment 5 indicated that there is indeed an effect of gender stereotypicality on speech processing, in the sense that stereotypical voices are processed faster than are nonstereotypical voices. There is, then, a processing deficit, visible in longer RTs, that is incurred from the nonstereotypical voices that we examined in this experiment.

In the following section, I describe the priming paradigm that we employed in our Experiment 6, which was a primed auditory naming task.

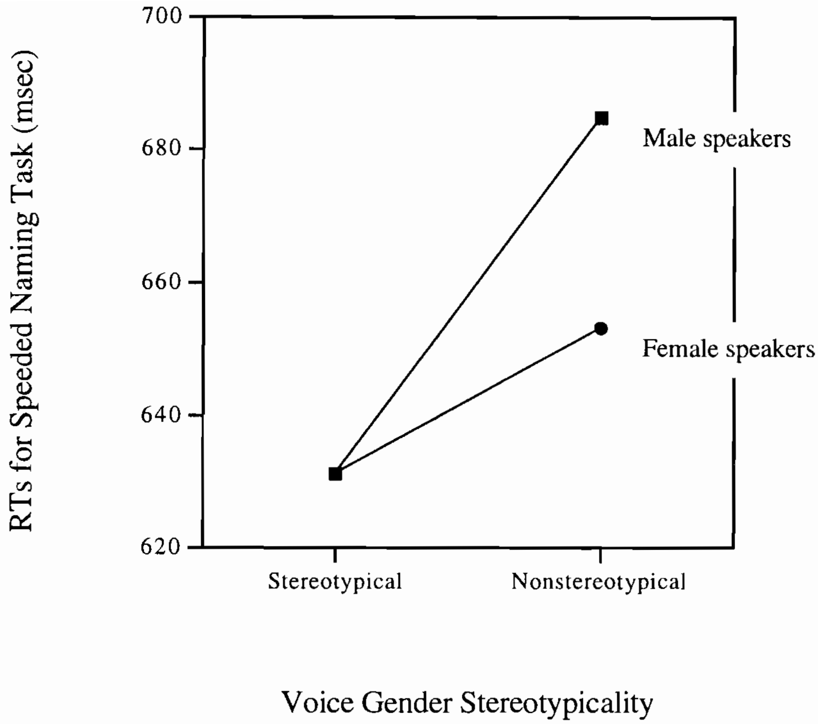

Figure 5.1: Reaction Time (RT) results for the four voices in Experiment 5, *Speeded Naming Task (Audio-Only).* Time is given in milliseconds along the y-axis, and voice gender stereotypicality is given along the x-axis. RTs are averaged across all participants for all 48 words as spoken by each speaker. Female speakers are indicated by filled circles, and male speakers are indicated by filled squares. RTs for stereotypical speakers are plotted at the left of the figure, and RTs for nonstereotypical speakers are plotted at the right of the figure. The average RTs for each voice are as follows: Stereotypical Male Speaker = 631.26 msec; Stereotypical Female Speaker = 631.30 msec; Nonstereotypical Male Speaker = 684.79 msec; and Nonstereotypical Female Speaker = 653.34 msec.

5.2 THE PRIMING PARADIGM

The procedure that we used in Experiment 6 to investigate the effects of facially-induced gender stereotypes on speech processing is a version of a commonly used priming paradigm (e.g., Posner & Snyder 1975, Rosch 1978). The general form of the paradigm is to present, in each of a series of trials, a "target" stimulus, about which some decision is to be made, and to precede that stimulus by either a "prime" or a neutral control item. Although the prime may be related to the target that it precedes, the decision is to be made on the basis of that target alone. Performance in a no-prime condition of the task is used to establish a baseline value of response time. The RT for the primed condition is compared to this baseline, and deviations are assumed to represent the influence of the prime on the processing of the target stimulus.

If a preliminary prime activates information in common with that processed from the target stimulus, RT should be reduced below the no-prime baseline, even though the prime is not directly relevant to the decision. If the prime activates information that mismatches the information activated by the target stimulus, two outcomes are possible. First, there may be no effect at all. Alternatively, responses may be slowed, in comparison to the baseline RT. The latter effect may occur solely because the prime draws processing capacity away from the target stimulus, or because it arouses information that is incongruent with—not just irrelevant to—processing of the target stimulus.

The magnitude of priming effects (i.e., the differences in RT to target stimuli when preceded versus *not* preceded by a prime) has been found to depend in part on two variables. One is the interval between the onsets of the prime and target stimulus (prime-to-stimulus interval, or PSI). If it takes time for a prime to activate information in memory, then as the PSI increases, there should be more opportunity for prime-produced activation to occur before the target decision is made, and thus more influence of the prime on responses. As would be expected on this basis, the magnitude of priming effects has been found to increase as the PSI increases, over a limited range (e.g., 0 msec to approximately 500 msec, in Posner & Snyder 1975; Rosch 1975). The minimum interval at which an effect is found is indicative of how quickly a prime can activate information in memory that is relevant to the target decision, since below that interval, the decision is apparently reached before significant activation occurs.

A second variable that influences the magnitude of priming effects is the nature of the judgment to be made about the target stimulus. A variety of tasks have been used in conjunction with the priming paradigm, including judgments as to whether strings of letters form words (Neely 1977), whether two letters are physically identical (Posner & Snyder 1975), and whether two objects are members of a common superordinate category (Rosch 1975). These tasks can be viewed as varying both in the nature of the information that is used in the decision about the target stimulus, and, concurrently, in the internal processes that must be performed in order to extract that information and reach the decision. For example, the decision as to whether two objects are physically identical is thought to rely on a non-illusory stimulus representation that is achieved early in the process of visual encoding, whereas the decision as to whether two visually distinct objects are from the same abstract superordinate category would require not only visual encoding, but also retrieval of information in memory regarding category membership.

In general, the nature of the target judgment constrains just how abstract primed information can be and still affect the response, because it determines how much processing of the target stimulus is performed before the response decision is made. If the prime activates information of a sort that becomes available from the stimulus only after further processing, any match or mismatch between the two sources of information will become evident too late, after the decision has been made. For example, Rosch (1975) found that a relatively abstract prime word like "furniture" had no effect on the task of deciding whether two objects in the primed class were physically identical, but did influence decisions as to whether they came from a common superordinate category. As Rosch interpreted it, the superordinate prime activated information at a more abstract level than that used in the decision as to physical identity.

In the following section, I discuss our primed auditory naming task experiment.

5.3 EXPERIMENT 6: FACE-PRIMED SPEEDED NAMING TASK

5.3.1 Experiment 6 Purpose

This experiment examines the perception of speech tokens that are primed by a visual representation of the speaker. Lemm, Dabady, Banaji, and Kluewer (1999) explored how pictorial primes activate gender stereotypes, employing a task similar to that used by Banaji and Hardin (1996), as I described earlier in §5.0. Lemm and colleagues showed that male and female facial photographs evoke strong gender primes which then affect ensuing processing in a word judgment task. We employed this methodology to examine the effects of face primes on speech perception.

In this experiment, a visual prime (in the form of a still shot of one of the four faces identified in Chapter 4) was shown before each audio token was presented to see how the integration of the face prime with the voice would influence repetition reaction times. Primes and audio tokens were presented in various combinations of gender stereotypicality.

The methods for this experiment were identical to those of the prior naming task, except that each audio token was preceded by a visual representation of the face of the speaker. Again, participants were told to name the word aloud as quickly as possible after hearing each target word.

We predicted that we would find an interaction between voice and face stereotypicality, so that audiovisual tokens with faces and voices that are matched for stereotypicality will be processed more quickly than those that are unmatched for stereotypicality, and stereotypical voices will be processed more quickly in general than nonstereotypical voices.

5.3.2 Experiment 6 Methods

5.3.2.1 Participants

The participants for Experiment 6 were 25 Ohio State undergraduate and graduate students (17 females, 8 males; 21 undergraduates, 4 graduates). Some of the participants received partial course credit in their undergraduate linguistics courses in exchange for their participation, with the remaining participants receiving $5.00 for their participation. All were native speakers of American English, and all completed personal history questionnaires before taking part in the experimental session. None reported any speech or hearing disorders. All participants were debriefed immediately after their session.

5.3.2.2 Stimuli

The stimuli for Experiment 6 were audiovisual, consisting of two parts: audio speech tokens, and pictures of faces. The audio stimuli for Experiment 6 were the same 192 sound files of the 48 naturally-produced English words, each uttered by the four selected speakers, that were used in Experiment 5 as described above in §5.1.2.2. (See Appendix F for a listing of these words.) The visual stimuli for Experiment 6 were the four picture files of the two female and two male faces that were selected for stimuli creation in §4.5 of the previous chapter (Female Face 3, Female Face 9, Male Face 2, and Male Face 3). (See Appendix C for a presentation of these faces.)

The audio and visual stimuli were mixed and matched so that each speaker's voice was paired with each face 12 times throughout the series of 192 tokens.

5.3.2.3 Procedure

Before starting Experiment 6, each participant read a printed list of experiment instructions in which they were told that they would be hearing a series of voices uttering different words as they saw pictures of the speakers' faces, and that their task would be to repeat back the word that they heard as quickly and accurately as possible.

Participants were tested in a sound-attenuated booth. They were fitted with the same Sennheiser HMD 410 Headset/Boom Microphone Set that was used in Experiment 5.

MEL Professional software on a Dell 466/M computer was used to present the stimuli sets to participants and collect their responses. In each trial, participants were first presented with one of the visual face stimuli, positioned in the middle of the computer screen with a black background. After a prime-to-stimulus interval (PSI) of 250 msec, the participants heard one of the audio stimuli binaurally via the Sennheiser HMD 410 set's headphones, and repeated the target word back into the set's microphone as soon as they could identify it. The participants' input into the microphone was first fed into a TTE 411AFS Amplifier for preamplification as before. The output from the amplifier was then fed into the microphone input jack of a 5-button Psychology Software Tools serial response box, which activated the box's sound-triggered timing circuit and caused the participant's response time to be logged in a computer data file.

Immediately after each response was logged, participants were presented with a visual display of their reaction time, in seconds. If participants failed to respond within 4 seconds, an indicator of "no response given" was logged. After the RT feedback was shown, an intertrial interval of 1000 msec occurred before the next stimuli set was presented.

Participants first heard a practice block of 20 stimuli to familiarize them with the procedure for the face-primed speeded naming task. All 192 of the audiovisual stimuli sets were then presented in random order to each participant, in 4 blocks of 48 stimuli each. Response latencies were measured from the start of the sound file (which corresponds with the onset of energy of the initial speech sound in the token word) until the participant's vocal response triggered the timing circuit in the response box. Each experimental session lasted approximately 20 minutes.

5.3.3 Experiment 6 Results and Discussion

The reaction time results for the Face-Primed Speeded Naming Task are given in Table 5.1 for each speaker—face combination. Results for the two female voices are graphically presented in Figure 5.2, and results for the two male voices are graphically presented in Figure 5.3. These results are averaged across the responses of all participants.

RT (msec)	Speaker	Face
585.61	Stereotypical Male Voice	Stereotypical Male Face
580.81	Stereotypical Male Voice	Nonstereotypical Male Face
593.16	Stereotypical Male Voice	Nonstereotypical Female Face
598.19	Stereotypical Male Voice	Stereotypical Female Face
630.31	Nonstereotypical Male Voice	Stereotypical Male Face
649.83	Nonstereotypical Male Voice	Nonstereotypical Male Face
628.21	Nonstereotypical Male Voice	Nonstereotypical Female Face
637.78	Nonstereotypical Male Voice	Stereotypical Female Face
573.81	Stereotypical Female Voice	Stereotypical Male Face
601.05	Stereotypical Female Voice	Nonstereotypical Male Face
581.05	Stereotypical Female Voice	Nonstereotypical Female Face
561.06	Stereotypical Female Voice	Stereotypical Female Face
587.10	Nonstereotypical Female Voice	Stereotypical Male Face
622.10	Nonstereotypical Female Voice	Nonstereotypical Male Face
606.67	Nonstereotypical Female Voice	Nonstereotypical Female Face
581.49	Nonstereotypical Female Voice	Stereotypical Female Face

Table 5.1: Reaction Time (RT) results for Experiment 6, *Face-Primed Speeded Naming Task*. Results for the four voices (stereotypical female and male, nonstereotypical female and male) are separated in the table, with RTs given in milliseconds in the first column of the table. Reaction time results are averaged across all participants' reactions to all 48 stimulus words for each speaker—face combination.

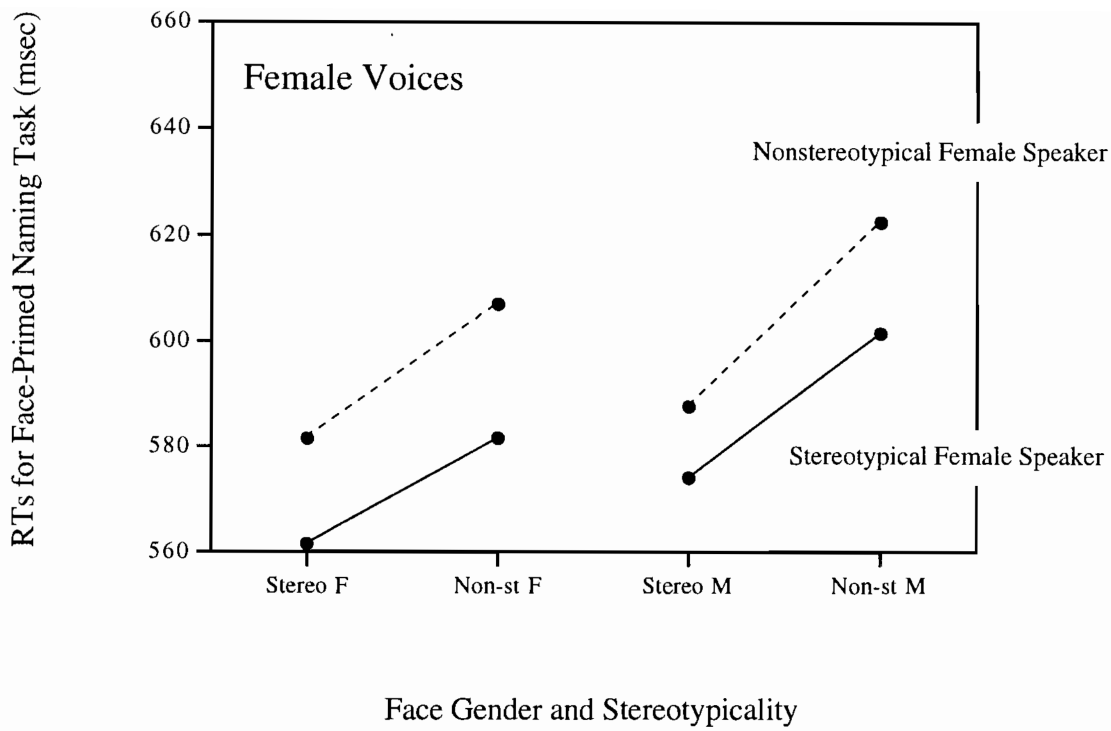

Figure 5.2: Reaction Time (RT) results for the Female voices in Experiment 6, *Face-Primed Speeded Naming Task*, as primed by each of the four faces. Time is given in milliseconds along the vertical axis, and face gender and stereotypicality are given along the horizontal axis. RTs are averaged across all participants for all 48 words. Results for the stereotypical female speaker are indicated by the plot points that are connected with a solid line, and results for the nonstereotypical female speaker are indicated by the plot points that are connected with a dashed line. Values for each data point are given in Table 5.1.

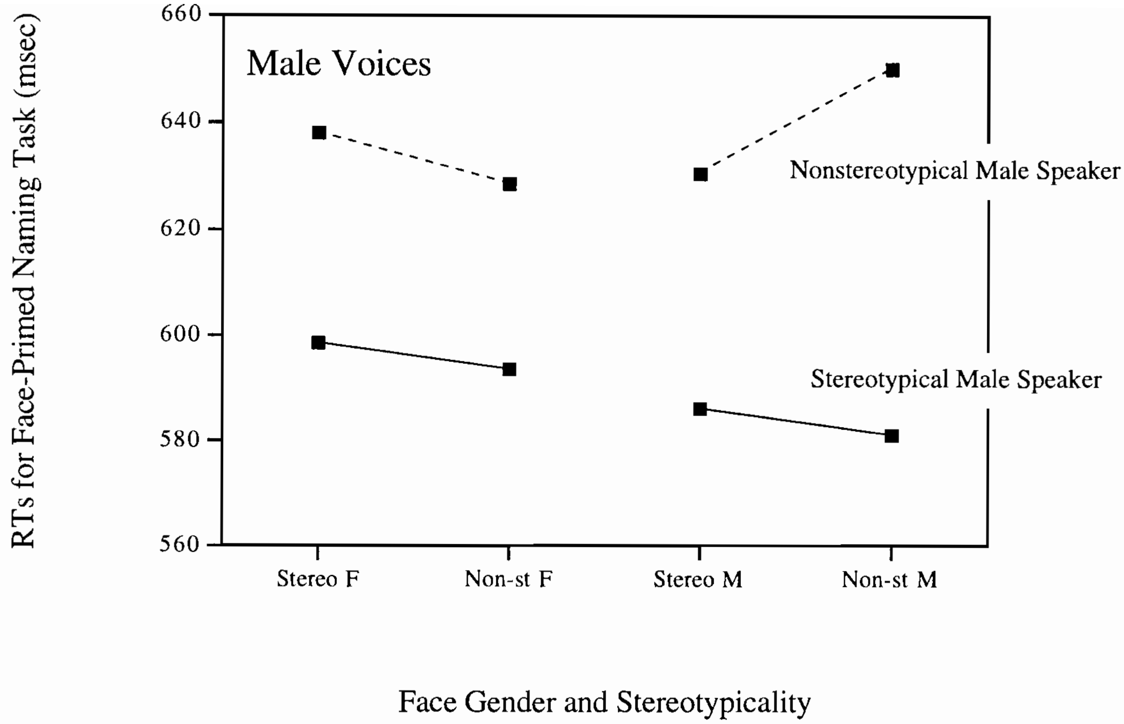

Figure 5.3: Reaction Time (RT) results for the Male voices in Experiment 6, *Face-Primed Speeded Naming Task*, as primed by each of the four faces. Time is given in milliseconds along the vertical axis, and face gender and stereotypicality are given along the horizontal axis. RTs are averaged across all participants for all 48 words. Results for the stereotypical male speaker are indicated by the plot points that are connected with a solid line, and results for the nonstereotypical male speaker are indicated by the plot points that are connected with a dashed line. Values for each data point are given in Table 5.1.

5.3.3.1 Results for the Female Voices

Preliminary analysis of the data for the male and female voices indicated that there was much variation in the way that data for each group patterned. This fact led us to conduct separate analyses of covariance on each voice gender category, to avoid muddling our final results.

We first conducted an ANCOVA on the RT data for the female voices, with "word duration" as the covariate and including three factors for analysis. The first factor was "voice stereotypicality" (stereotypical speaker or nonstereotypical speaker), the second factor was "face gender" (female face or male face), and the third factor was "face stereotypicality" (female face or male face).

Two significant main effects and one significant interaction were obtained for the female voices as the results of the ANCOVA. First, there was a significant main effect of voice stereotypicality [F (1,24) = 8.05; $p<.01$], with the stereotypical female voice inducing a significantly shorter RT than the nonstereotypical female voice. The second significant effect that we identified was the main effect of face stereotypicality [F (1,24) = 26.4; $p<.001$]. This indicates that female voices that were primed by stereotypical faces were responded to significantly faster than when the same voices were primed with nonstereotypical faces. The factor "face gender" was not significant.

A significant interaction of face gender and voice stereotypicality was also identified [F (1,24) = 5.7; $p<.05$]. This interaction indicates that the effects of voice stereotypicality and face gender are interdependent. This interaction is likely at least partly explained by the nature of the specific audio and visual stimuli that are in question here. Keeping in mind that the effect of this interaction is a relatively small, it may mostly be reflecting the effect of the nonstereotypical male face, for which the average gender identification RT in Experiment 4 was very slow, as well as the effect of the nonstereotypical female voice, for which the gender judgment accuracy calculated in Experiment 3 was quite low.

The results for the female voices in Experiment 6, then, show that there is again a significant effect of voice stereotypicality, replicating the result that we found in Experiment 5. In terms of the effects of the face primes on the processing of the female voices, there was also a significant main effects of face stereotypicality, indicating that stereotypical faces were faster to be processed overall, even if face–voice gender in the audiovisual token was not matched. Our expectations about how the gendered facial information would interact with speech processing were therefore confirmed for the female voices.

5.3.3.2 Results for the Male Voices

In terms of the male voices, however, our results are very different. As we did for the female voices, we conducted a by-items ANCOVA on the RT data for the male voices, with "word duration" as the covariate and the same three factors that we used to assess the primed female voices: "voice stereotypicality" (stereotypical speaker or nonstereotypical speaker), "face gender" (female face or male face), and "face stereotypicality" (female face or male face).

As we had found for the female voices, there was a very significant main effect of voice stereotypicality for the male voices [$F(1,24) = 25.6$; $p<.001$], replicating the voice stereotypicality effect that we found in Experiment 5.

Contra our findings for the female voices, however, in the case of the male voices there were no significant face effects, neither of either face gender nor of face stereotypicality. This means that there was no reliable face priming effect on the male voices.

5.3.3.3 General Priming Results

We concluded from the results discussed in the sections above that gender priming effects operate differently on the female voices and male voices that were used in our study. The fact that the face prime and target stimulus were presented within the 500 msec PSI "magic window," and elicited significant results, indicates that for the female voices, the face prime did indeed activate information in common with that which was processed from the target stimulus (as I discussed above in §5.2), in this case some feature of gender. So for the female voices, the effect of face stereotypicality is a clear priming effect.

It is not clear why the male voices seem so insulated from the face effects, but there could be a number of factors at work here. Perhaps this is due at least in part to variation in processing strategies for male and female voices in general; a number of researchers, for example, have claimed that perceivers employ different strategies when processing male versus female voices (see Kreiman 1997). Or perhaps there is an effect of the stricter culturally-mediated gender expectations that we have for features of male performance in this culture, as I mentioned in earlier chapters.

Or, another alternative is that this lack-of-effect could be due to something specifically inherent to the two male voices that we used in the present experiments. Recalling the gender judgment accuracy results that I gave in Chapter 4, we saw much more variability in accuracy for judging gender in the female voices, with the male voices as a group showing gender judgment accuracy that was much closer to ceiling level. This indicates that the male voices in our set were more clearly identifiable as "male" in general than were the female voices in our set. Therefore, the voice gender cues were stronger for the two male voices (versus the female voices) used in the experiments in this chapter, which might make it easier for the processing system to simply inhibit the face gender cue in the perception of the male voices. That is, perhaps the gender stereotype induced by the face is harder to integrate with the stronger voice gender cues of the male voices once the voices start.

5.4 SUMMARY

The results from Experiments 5 and 6 suggest that gender stereotypes about speakers are indeed used processing speech information. Further, gender stereotypes about speakers can be accessed by face information as well as voice information, and subsequently employed in processing. The significant voice stereotypicality and face effects in facilitating auditory naming imply that gender categories and word/phonological categories are inextricably integrated in speech perception. Additionally, gender information that influences speech perception can come from multiple modalities. Based on results such as these, it is clear that models of speech perception will necessarily need to be able to account for such interactions of multimodal, social and phonological categories.

CHAPTER 6

CONCLUSIONS

In this dissertation, I have presented research which suggests that gender stereotypes do indeed influence speech perception, even at very low levels of cognitive processing. I have used the techniques of multidimensional scaling and speeded category classification to assess gender stereotypes of faces and voices in an implicit fashion. I then used the faces and voices that were identified on the basis of these techniques to construct stimuli for use in a non-primed auditory naming task in which participants repeated words that were uttered by voices of varying stereotypicality. I employed the same auditory stimuli in a second task, a primed auditory naming task in which the primes were pictures of faces of assumed speakers.

My results from these two experiments indicate that gender stereotypes do indeed influence perception: in the non-primed auditory naming task, we found a significant main effect of voice stereotypicality on the processing and repetition of words. Stereotypical voices are faster to be processed. In the face-primed auditory naming task, we found that for FEMALE voices, there were significant effects of face stereotypicality on the speed with which the words in the task could be processed and repeated; face primes that were stereotypical induced faster processing times. These results indicate that information in the prime that is congruent with information in the target facilitates processing of the target.

In the case of the MALE voices, however, we found no significant effect of the faces on processing. We hypothesized that this could be due to a number of factors, including the likely fact that the male voices are particularly strong categories (recalling the high levels of gender judgment accuracy that were found) and were therefore more resistant to the influence of face-gender information, than were the female voices.

The work that I have presented in the previous chapters impinges upon current notions of phonological categories, as well as their representation. Phonological categories are generally understood to represent the basis of opposition. For example, in American English, the category /t/ is opposed to /p/ since contrasts in meaning may be based on this contrast, as seen in “tool” versus “pool.”

It is also generally understood that the manifestation of phonological categories is context-sensitive. For example, the manifestation of the category /t/ varies among “tool” [tʰ], “stool” [t], “loot” [t̚], and “atmosphere” [ʔ]. The pattern illustrated by these examples is robust, and the variation is predictable.

Autosegmental properties such as pitch are also generally understood to be context-dependent. In a language with contrastive tones (say high (H) and low (L)), the fundamental frequency of a H is higher at the beginning of an utterance than it is at the end. It is also accepted that a H produced by a female speaker is generally higher than that produced by a male speaker.

The research that I have discussed throughout this work supports a strong claim, however, that the gender of a speaker is a contextual cue for “segmental” categories as well as prosodic categories. Recall the audiovisual studies that I described in Chapter 3, which demonstrated that a sound was perceived as [s] when paired with a male face, but as [ʃ] when paired with a female face. Similar results were found for the categories [ʌ] and [ʊ].

This strongly suggests that other phonetic categories may be susceptible to this "gender effect," beyond those phonetic categories that are explicitly socially diagnostic for a given culture.

The implications of this work for phonological categories is that the notion of context must be expanded to include the gender of the speaker. A native speaker's competence includes the knowledge of how a given phonological category is manifested not only with respect to its position in an utterance and its relation to surrounding sounds, but also with respect to the gender of the speaker, as well as to any number of other culturally-relevant, socially-mediated factors. That is, part of being a native speaker of a language is the expectation that speakers of different genders will produce the same sound differently, and these expectations become coded in the representation of speech categories, and utilized by listeners in speech perception processing.

APPENDIX A

The following 24 words were included as stimuli in Experiments 1 and 3, as described in Chapter 4.

bath	life
boat	lot
case	mile
choice	mood
death	path
dog	sheet
food	shirt
girl	term
hill	town
home	voice
jazz	wall
kid	wish

APPENDIX B

	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	m1	m2	m3	m4	m5	m6	m7	m8	m9	m10
f1	.	3.72	4.89	3.7	3.85	4.04	4.33	3.83	4.28	4.13	5.98	5.72	6.3	6.65	6.33	6.54	6.07	7.5	6.24	6.24
f2	3.72	.	3.17	3.35	2.72	3.33	3.74	2.98	3.87	3.26	5.63	5.91	5.57	6.02	6.63	6.15	5.22	6.83	5.76	6.59
f3	4.89	3.17	.	3.61	3.35	3.33	4.39	2.87	3.15	3.04	5.98	5.76	6.17	7.28	5.96	6.13	5.7	6.52	6.02	6.59
f4	3.7	3.35	3.61	.	3.22	3.26	4.7	3.96	4.57	2.98	6.02	6.54	6.54	6.98	6.8	6.63	5.5	7.83	6.59	7
f5	3.85	2.72	3.35	3.22	.	2.48	3.89	3.09	3.52	3.13	6.26	6.04	5.96	6.74	6.54	6.02	5.63	7.35	6.22	6.59
f6	4.04	3.33	3.33	3.26	2.48	.	4.3	3.22	4.09	2.61	6.59	5.63	6.52	6.93	6.74	6.41	6.07	7.26	6.37	6.61
f7	4.33	3.74	4.39	4.7	3.89	4.3	.	4.41	3.76	4.83	5.48	6.15	5.35	5.76	5.96	5.5	5.74	6.93	5.83	6.35
f8	3.83	2.98	2.87	3.96	3.09	3.22	4.41	.	3.04	2.48	6.3	5.7	6.17	6.63	6.46	5.87	5.85	7.09	5.7	6.72
f9	4.28	3.87	3.15	4.57	3.52	4.09	3.76	3.04	.	4.13	6.35	5.89	5.93	6.3	6.04	5.93	5.54	7.26	5.63	6.15
10	4.13	3.26	3.04	2.98	3.13	2.61	4.83	2.48	4.13	.	5.8	6.07	6.11	6.78	6.5	6.46	5.74	6.54	6.43	6.54
m1	5.98	5.63	5.98	6.02	6.26	6.59	5.48	6.3	6.35	5.8	.	3.28	3.04	4.22	3.35	3.48	3.22	4.52	3.54	3
m2	5.72	5.91	5.76	6.54	6.04	5.63	6.15	5.7	5.89	6.07	3.28	.	3.93	4.7	4.33	4.11	3.46	5.22	3.46	4.43
m3	6.3	5.57	6.17	6.54	5.96	6.52	5.35	6.17	5.93	6.11	3.04	3.93	.	3.3	3.37	3.09	3.13	4.54	4.11	3.91
m4	6.65	6.02	7.28	6.98	6.74	6.93	5.76	6.63	6.3	6.78	4.22	4.7	3.3	.	4.11	2.78	4.98	4.39	4.96	4.09
m5	6.33	6.63	5.96	6.8	6.54	6.74	5.96	6.46	6.04	6.5	3.35	4.33	3.37	4.11	.	3.83	4.54	2.89	4.13	4.07
m6	6.54	6.15	6.13	6.63	6.02	6.41	5.5	5.87	5.93	6.46	3.48	4.11	3.09	2.78	3.83	.	3.91	4.59	4.5	3.24
m7	6.07	5.22	5.7	5.5	5.63	6.07	5.74	5.85	5.54	5.74	3.22	3.46	3.13	4.98	4.54	3.91	.	5.11	4.07	3.61
m8	7.5	6.83	6.52	7.83	7.35	7.26	6.93	7.09	7.26	6.54	4.52	5.22	4.54	4.39	2.89	4.59	5.11	.	4.96	5.07
m9	6.24	5.76	6.02	6.59	6.22	6.37	5.83	5.7	5.63	6.43	3.54	3.46	4.11	4.96	4.13	4.5	4.07	4.96	.	4.43
10	6.24	6.59	6.59	7	6.59	6.61	6.35	6.72	6.15	6.54	3	4.43	3.91	4.09	4.07	3.24	3.61	5.07	4.43	.

Table B.1: Matrix of similarity judgment values for pairs of the voice stimuli that were rated in Experiment 1. This matrix was the input for the multidimensional scaling solution for voice similarity. Each cell in the matrix represents the averaged similarity rating for the voice pair, as identified by the row and column labels (e.g., "f1" represents Female Speaker 1, "m1" represents Male Speaker 1, etc.). Similarity rating values are averaged across all subjects for each pair of voices.

APPENDIX C

These 20 faces were included as stimuli in Experiments 2, 4, and 6. The numbers of the faces correspond to the data point numbers given in the multidimensional scaling solution for faces, which is displayed in Figures 4.6 and 4.7.

Figure C.1: Female Face 1

Figure C.2: Female Face 2

Figure C.3: Female Face 3

Figure C.4: Female Face 4

(continued)

Figure C.5: Female Face 5

Figure C.6: Female Face 6

Figure C.7: Female Face 7

Figure C.8: Female Face 8

Figure C.9: Female Face 9

Figure C.10: Female Face 10

(continued)

Figure C.11: Male Face 1

Figure C.12: Male Face 2

Figure C.13: Male Face 3

Figure C.14: Male Face 4

Figure C.15: Male Face 5

Figure C.16: Male Face 6

(continued)

Figure C.17: Male Face 7

Figure C.18: Male Face 8

Figure C.19: Male Face 9

Figure C.20: Male Face 10

APPENDIX D

	f1	f1	f3	f4	f5	f6	f7	f8	f9	f10	m1	m2	m3	m4	m5	m6	m7	m8	m9	m10
f1	.	6.5	6.04	5.42	7.58	7.42	6.38	6.88	7.17	6.04	7.75	7.71	6.79	7.92	8.13	7.88	8.79	8.17	8.08	8
f2	6.5	.	4.92	4.83	6.63	6.38	4.79	5.71	3.58	6.92	5.58	5.88	4.58	6.79	6.38	8.04	8.5	7.25	7.38	4.83
f3	6.04	4.92	.	3.92	6.25	5.17	3.92	3.54	4.04	6.71	5.25	6.79	4.71	7.92	6.04	8.25	7.79	6.83	7.67	5.04
f4	5.42	4.83	3.92	.	6.25	5.08	3.5	5.21	4.33	6.83	6.08	6.58	5.96	7.13	5.75	7.67	7.88	7.63	6.71	5.67
f5	7.58	6.63	6.25	6.25	.	6.13	6.21	5.29	6.58	5.17	5.46	5.38	6.46	6.17	5.25	7.92	7.71	7.13	4.38	7.38
f6	7.42	6.38	5.17	5.08	6.13	.	4.08	4.79	3.71	7.04	5.04	6.79	3.92	7.71	6.63	8.33	8.33	5.13	7	4.71
f7	6.38	4.79	3.92	3.5	6.21	4.08	.	4.67	3.79	6.33	4.33	6.88	4.46	6.71	4	7.88	7.63	5.17	5.67	5.21
f8	6.88	5.71	3.54	5.21	5.29	4.79	4.67	.	4.5	5.33	4.46	6.04	6	7.25	4.96	8.08	7.88	5.42	6.5	5.38
f9	7.17	3.58	4.04	4.33	6.58	3.71	3.79	4.5	.	6.88	4.42	7.17	3.71	7.08	5.17	8	8.21	5.21	6.54	4.92
10	6.04	6.92	6.71	6.83	5.17	7.04	6.33	5.33	6.88	.	7.21	5.88	5.38	5.04	5.08	6.96	7.63	7.88	5.25	7.75
m1	7.75	5.58	5.25	6.08	5.46	5.04	4.33	4.46	4.42	7.21	.	4.79	4.21	6.04	2.5	7.75	6.75	4.71	4.17	4.79
m2	7.71	5.88	6.79	6.58	5.38	6.79	6.88	6.04	7.17	5.88	4.79	.	5.79	6.38	3.79	8.04	8	4.75	5.17	6.38
m3	6.79	4.58	4.71	5.96	6.46	3.92	4.46	6	3.71	5.38	4.21	5.79	.	5.58	4.83	7.17	7.17	4.25	5.33	4.29
m4	7.92	6.79	7.92	7.13	6.17	7.71	6.71	7.25	7.08	5.04	6.04	6.38	5.58	.	4.25	5.21	6.17	7.08	3.92	7
m5	8.13	6.38	6.04	5.75	5.25	6.63	4	4.96	5.17	5.08	2.5	3.79	4.83	4.25	.	6.5	6.83	5.33	2.92	5.42
m6	7.88	8.04	8.25	7.67	7.92	8.33	7.88	8.08	8	6.96	7.75	8.04	7.17	5.21	6.5	.	4.25	8.13	6.21	7.83
m7	8.79	8.5	7.79	7.88	7.71	8.33	7.63	7.88	8.21	7.63	6.75	8	7.17	6.17	6.83	4.25	.	7.46	6.71	7.71
m8	8.17	7.25	6.83	7.63	7.13	5.13	5.17	5.42	5.21	7.88	4.71	4.75	4.25	7.08	5.33	8.13	7.46	.	7	5.13
m9	8.08	7.38	7.67	6.71	4.38	7	5.67	6.5	6.54	5.25	4.17	5.17	5.33	3.92	2.92	6.21	6.71	7	.	6
10	8	4.83	5.04	5.67	7.38	4.71	5.21	5.38	4.92	7.75	4.79	6.38	4.29	7	5.42	7.83	7.71	5.13	6	.

Table D.1: Matrix of similarity judgment values for pairs of the face stimuli that were rated in Experiment 1. This matrix was the input for the multidimensional scaling solution for face similarity. Each cell in the matrix represents the averaged similarity rating for the face pair, as identified by the row and column labels (e.g., "f1" represents Female Face 1, "m1" represents Male Face 1, etc.). Similarity rating values are averaged across all subjects for each pair of faces.

APPENDIX E

The following formulae, adapted from Edwards (1985:255-256), can be used to calculate new Dimension 1 (x) and Dimension 2 (y) coordinates for a two-dimensional map space when it is desirable to rotate the axes given by the original multidimensional solution.

For any point with the coordinates (x, y) in the 2-dimensional space, find (x′, y′) in the rotated space.

1. Calculate the constants D and ϕ (D is the length of the hypotenuse of a triangle with its corners at points (0, 0), (x, y), and the x-value on the x-axis; and ϕ is the degree of the angle that radiates from (0, 0) in that triangle).

   $$D = \sqrt{x^2 + y^2}$$

   $$\cos \phi = x \div D$$

   $$\sin \phi = y \div D$$

2. Determine the constant θ (this is the degree of the angle of rotation of the set of data points that is necessary to align the new axes as desired — measure the degree of this angle with a protractor).

(continued)

3. Calculate ϕ' (in the rotated solution, ϕ' will be the new degree of the angle that radiates from (0, 0) in the new triangle with corners at (x', y') and x' on the x'-axis).

$$\phi' = \phi + \theta$$

$$\cos \phi' = (\cos \phi \times \cos \theta) - (\sin \phi \times \sin \theta)$$

$$\sin \phi' = (\sin \phi \times \cos \theta) + (\cos \phi \times \sin \theta)$$

4. Calculate the new x' and y' values in the rotated solution.

$$x' = \cos \phi' \times D$$

$$y' = \sin \phi' \times D$$

APPENDIX F

The following 48 words were included as stimuli in Experiments 5 and 6, as described in Chapter 5.

bath	fig	king	shirt
beach	foam	lake	sin
bird	food	leg	song
boat	girl	life	term
case	goal	lot	town
chin	hat	mile	tube
choice	height	mood	van
coal	hill	mud	voice
dawn	home	noise	wall
death	jazz	path	wine
dirt	job	pope	wish
dog	kid	sheet	youth

BIBLIOGRAPHY

Adorno, Theodore W., Else Frenkel-Brunswik, Daniel J. Levinson, & R. Nevitt Sanford. (1950) *The Authoritarian Personality.* New York: Harper.

Allport, Gordon W. (1954) *The Nature of Prejudice.* Reading, Massachusetts: Addison-Wesley.

Aronson, Elliot. (1995) *The Social Animal* (Seventh Edition). New York: W.H. Freeman and Company.

Ashmore, Richard D., Frances K. Del Boca, & Arthur J. Wohlers. (1986) Gender stereotypes. In Richard D. Ashmore & Frances K. Del Boca (Eds.), *The Social Psychology of Male-Female Relations* (pp. 69-119). Orlando, Florida: Academic Press.

Banaji, Mahzarin R. & Anthony G. Greenwald. (1995) Implicit stereotyping in false fame judgments. *Journal of Personality and Social Psychology, 68,* 181-198.

Banaji, Mahzarin R. & Curtis D. Hardin. (1996) Automatic stereotyping. *Psychological Science, 7:3,* 136-141.

Banaji, Mahzarin R., Curtis D. Hardin, & Alexander J. Rothman. (1993) Implicit stereotyping in person judgment. *Journal of Personality and Social Psychology, 65,* 272-281.

Baugh, John. (1999) Controlled recordings of John Baugh used in experimental studies of housing discrimination based on speech [On-line]. Available: http://www.stanford.edu/~jbaugh/baugh.fft

Beall, Anne E. & Robert J. Sternberg. (1993) *The Psychology of Gender.* New York: Guilford Press.

Beebe, Leslie M. (1981) Social and situational factors affecting communicative strategy of dialect code-switching. *International Journal of the Sociology of Language, 32,* 139-149.

Bem, Sandra. (1993) *The Lenses of Gender: Transforming the Debate on Sexual Inequality.* New Haven: Yale University Press.

Bem, Sandra Lipsitz. (1981) Gender schema theory: A cognitive account of sex typing. *Psychological Review, 88:4,* 354-364.

Bem, Sandra Lipsitz. (1974) The measurement of psychological androgeny. *Journal of Consulting and Clinical Psychology, 42,* 155-162.

Berry, Dianne S. & Leslie McArthur. (1985) Some components and consequences of a babyface. *Journal of Personality and Social Psychology, 48:2,* 312–23.

Best, John B. (1999) *Cognitive Psychology* (Fifth Edition). Belmont, California: Wadsworth Publishing Company.

Bing, Janet M. & Victoria L. Bergvall. (1996) The question of questions: Beyond binary thinking. In Victoria L. Bergvall, Janet M. Bing, & Alice F. Freed (Eds.), *Rethinking Language and Gender Research: Theory and Practice* (pp. 1-30). New York: Addison Wesley Longman.

Bladon, Anthony, Caroline Henton, & J.B. Pickering. (1984) Towards an auditory theory of speaker normalization. *Language & Communication, 4:1,* 59-69.

Brehm, Sharon S. & Saul M. Kassin. (1996) *Social Psychology* (3rd Edition). Boston: Houghton Mifflin Company.

Brewer, Marilyn B. (1994) The social psychology of prejudice: Getting it all together. In Mark P. Zanna & James M. Olson (Eds.), *The Psychology of Prejudice: The Ontario Symposium, Volume 7* (pp. 315-329). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Brewer, Marilyn B. (1988) A dual process model of impression formation. In Thomas K. Srull & Robert S. Wyer, Jr. (Eds.), *Advances in Social Cognition: Volume 1. A Dual Process Model of Impression Formation* (pp. 1-36). Hillsdale, NJ: Lawrence Erlbaum.

Brown, Elizabeth & David I. Perrett. (1993) What gives a face its gender? *Perception, 22,* 829-840.

Burton, A. Mike, Vicki Bruce, & Neal Dench. (1993) What's the difference between men and women? Evidence from facial measurement. *Perception, 22,* 153-176.

Busey, Thomas A. (1998) Physical and psychological representations of faces: Evidence from morphing. *Psychological Science, 9:6,* 476-483.

Butler, Judith. (1993) *Bodies that Matter: On the Discursive Limits of 'Sex.'* New York: Routledge.

Butler, Judith. (1990) *Gender Trouble: Feminism and the Subversion of Identity.* New York: Routledge.

Cameron, Deborah. (1997) Performing gender identity: Young men's talk and the construction of heterosexual masculinity. In Sally Johnson & Ulrike Hanna Meinhof (Eds.), *Language and Masculinity* (pp. 47-63). Oxford: Blackwell Publishers.

Chambers, John K. (1995) *Sociolinguistic Theory: Linguistic Variation and its Social Significance.* Oxford: Blackwell.

Chapman, L. John. (1967) Illusory correlation in observational report. *Journal of Verbal Learning and Verbal Behavior, 6,* 151-155.

Chideya, Farai. (1999) *The Color of Our Future.* New York: William Morrow and Company, Inc.

Chomsky, Noam & Morris Halle. (1968) *The Sound Pattern of English.* New York: Harper & Row.

Constantinople, Anne. (1973) Masculinity-femininity: An exception to a famous dictum? *Psychological Bulletin, 80:5,* 389-407.

Cross, Susan E. & Hazel Rose Markus. (1993) Gender in thought, belief, and action: A cognitive approach. In Anne E. Beall & Robert J. Sternberg (Eds.), *The Psychology of Gender* (pp. 55-98). New York: Guilford Press.

Daly, Nicola. (1997) *Aspects of Talker Gender Difference in Speechreading.* Ph.D. Dissertation, La Trobe University. Bundoora, Victoria, Australia.

Daly, Nicola, John Bench, & Hillary Chappell. (1996) Interpersonal impressions, gender stereotypes, and visual speech. *Journal of Language and Social Psychology, 15:4,* 468-479.

Deacon, Terrence W. (1997) *The Symbolic Species: The Co-Evolution of Language and the Brain.* New York: W.W. Norton & Company.

Deaux, Kay. (1995) How basic can you be? The evolution of research on gender stereotypes. *Journal of Social Issues, 51:1,* 11-20.

Deaux, Kay & Marianne LaFrance. (1998) Gender. In Daniel T. Gilbert, Susan T. Fiske, & Gardner Lindzey (Eds.), *The Handbook of Social Psychology* (Fourth Edition; pp.788-827). New York: Oxford University Press.

Deaux, Kay & Laurie L. Lewis. (1984) Structure of gender stereotypes: Interrelationships among components and gender label. *Journal of Personality and Social Psychology, 46,* 991-1004.

Deaux, Kay, Ward Winton, Maureen Crowley, & Laurie L. Lewis. (1985) Level of categorization and content of gender stereotypes. *Social Cognition, 3:2,* 145-167.

Devine, Patricia G. (1989) Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56:1,* 5-18.

Dovidio, John F., Nancy Evans, & Richard B. Tyler. (1986) Racial stereotypes: the contents of their cognitive representations. *Journal of Experimental Social Psychology, 22,* 22-37.

Driver Leinbach, Mary & Beverly I. Fagot. (1993) Categorical habituation to male and female faces: Gender schematic processing in infancy. *Infant Behavior and Development, 16,* 317-332.

Duckett, John & Laurence Baskin. (1993) Genitoplasty for intersex anomalies. *European Journal of Pediatrics, 152:Suppl 2,* S80-S84.

Eckert, Penelope & Sally McConnell-Ginet. (1999) New generalizations and explanations in language and gender research. *Language in Society, 28:2,* 185-202.

Edwards, Jan. (1985) *Mandibular Rotation and Translation During Speech.* Ph.D. Dissertation, The City University of New York. New York, New York.

Egan, James P. (1949) Articulation testing methods. *Laryngoscope, 58,* 955-991.

Eisenberg, Leon. (1995) The social construction of the human brain. *American Journal of Psychiatry, 152:11,* 1563-1575.

Epstein, Julia. (1990) Either/or-neither/both: Sexual ambiguity and the ideology of gender. *Genders, 7,* 99-142.

Faludi, Susan. (1991) *Backlash: the Undeclared War Against American Women.* New York: Doubleday.

Fasold, Ralph. (1987) *The Sociolinguistics of Society.* Oxford, England: Basil Blackwell Ltd.

Fausto-Sterling, Anne. (1992) *Myths of Gender: Biological Theories About Women and Men* (Second Edition). New York: BasicBooks.

Fiske, Susan T. (1998) Stereotyping, prejudice, and discrimination. In Daniel T. Gilbert, Susan T. Fiske, & Gardner Lindzey (Eds.), *The Handbook of Social Psychology* (Volume 2; pp. 357-410). New York: Oxford University Press.

Fiske, Susan T. & Susan E. Taylor. (1991) *Social Cognition* (2nd Edition). New York: McGraw-Hill.

Fodor, Jerry A. (1983) *The Modularity of Mind: An Essay on Faculty Psychology.* Cambridge, Massachusetts: MIT Press.

Freund, John E. & Frank J. Williams. (1966) *Dictionary/Outline of Basic Statistics.* London: Constable and Company, Ltd.

Funderburg, Lise. (1994) *Black, White, and Other: Biracial Americans Talk About Race and Identity.* New York: William Morrow and Comany, Inc.

Fuster-Duran, Angela. (1996) Perception of conflicting audio-visual speech: An examination across Spanish and German. In David G. Stork & Marcus E.

Hennecke (Eds.), *Speech Reading by Humans and Machines* (pp. 135-144). Berlin: Springer-Verlag.

Gaertner, Samuel L. & John P. McLaughlin. (1983) Racial stereotypes: Associations and ascriptions of positive and negative characteristics. *Social Psychology Quarterly, 46,* 23-40.

Gaudio, Rudolph. (1994) Sounding gay: Pitch properties in the speech of gay and straight men. *American Speech, 69:1,* 30-57.

Gelfer, Marylou Pausewang. (1993) A multidimensional scaling study of voice quality in females. *Phonetica, 50,* 15-27.

Giles, Howard, Nikolas Coupland, & Justine Coupland. (1991) Accommodation theory: Communication, context, and consequence. In Howard Giles, Justine Coupland, & Nikolas Coupland (Eds.), *Contexts of Accommodation: Developments in Applied Sociolinguistics* (pp. 1-68). Cambridge: Cambridge University Press.

Glenn, Donna. (1999) Board allows girl to wear pants at commencement. *The Columbus Dispatch, 19 May,* 1A-2A.

Glick, Peter & Susan T. Fiske. (1999) Gender, power dynamics, and social interaction. In Myra Marx Ferree, Judith Lorber, & Beth B. Hess (Eds.), *Revisioning Gender* (pp. 365-398). Thousand Oaks, California: Sage Publications.

Graddol, David & Joan Swann. (1989) *Gender Voices.* Oxford: Blackwell Publishers.

Grammer, Karl & Randy Thornhill. (1994) Human (*Homo sapiens*) Facial Attractiveness and Sexual Selection: The Role of Symmetry and Averageness. *Journal of Comparative Psychology, 108:3,* 233-242.

Grassegger, Hans. (1995) McGurk effect in German and Hungarian listeners. *Proceedings of the International Congress of Phonetic Sciences, Stockholm, 4:3,* 210-213.

Green, Raymond J. & Richard D. Ashmore. (1998) Taking and developing pictures in the head: Assessing the physical characteristics of eight gender types. *Journal of Applied Psychology, 28,* 1609-1636.

Hall, Kira & Veronica O'Donovan. (1996) Shifting gender positions among Hindi-speaking hijras. In Victoria L. Bergvall, Janet M. Bing, & Alice F. Freed (Eds.), *Rethinking Language and Gender Research: Theory and Practice*(pp. 228-266). New York: Addison Wesley Longman.

Hamilton, David L., Steven J. Sherman, & Catherine M. Ruvolo. (1990) Stereotype-based expectancies: Effects in information processing and social behavior. *Journal of Social Issues, 46:2,* 35-60.

Hansen, Christine H. & Ranald D. Hansen. (1988b) How rock music videos can change what's seen when boy meets girl: Priming stereotypic appraisal of social interactions. *Sex Roles, 19,* 287-316.

Henton, Caroline G. & R. Anthony W. Bladon. (1985) Breathiness in normal female speech: Inefficiency versus desirability. *Language & Communication, 5:3,* 221-227.

Herman, Rebecca. (1998) *Intonation and Discourse Structure in English: Phonological and Phonetic Markers of Local and Global Discourse Structure.* Ph.D. Dissertation, The Ohio State University. Columbus, Ohio.

Herman, Rebecca & Julia Tevis McGory. (1999) Mapping intonational transcribers' intuitive tone spaces. *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 2331-2334).

Hilton, James L. & William von Hippel. (1996) Stereotypes. *Annual Review of Psychology 1996, 47,* 237-271.

Holmes, Janet. (1999) *Language in Society: Special Issue: Communities of Practice in Language and Gender Research, 28:2.*

Holmes, Janet & Miriam Meyerhoff. (1999) The Community of Practice: Theories and methodologies in language and gender research. *Language in Society, 28:2,* 173-184.

Hudson, Richard A. (1980) *Sociolinguistics.* Cambridge: Cambridge University Press.

Huxley, Aldous. (1954) *The Doors of Perception.* New York: Harper.

Jacobs, Sue-Ellen & Jason Cromwell. (1992) Visions and revisions of reality: reflections on sex, sexuality, gender, and gender variance. *Journal of Homosexuality, 23:4,* 43-69.

Johnson, Keith. (1997) *Acoustic and Auditory Phonetics.* Cambridge, Massachusettes: Blackwell Publishers Ltd.

Johnson, Keith. (1990) The role of perceived speaker identity in *F0* normalization of vowels. *Journal of the Acoustical Society of America, 88:2,* 642-654.

Johnson, Keith, Elizabeth A. Strand, & Mariapaola D'Imperio. (1999) Auditory-visual integration of talker gender in vowel perception. *Journal of Phonetics, 27:4,* 359-384.

Jones, Edward E. (1998) Major developments in five decades of social psychology. In Daniel T. Gilbert, Susan T. Fiske, & Gardner Lindzey (Eds.), *The Handbook of Social Psychology* (Fourth Edition; pp. 3-57). New York: Oxford University Press.

Klatt, Dennis H. (1986) The problem of variability in speech recognition and in models of speech perception. In Joseph S. Perkell & Dennis H. Klatt (Eds.), *Invariance and Variability in Speech Processes* (pp. 300-319). Hillsdale, New Jersey: Erlbaum.

Klatzky, Roberta, Gale L. Martin, & Robbie A. Kane. (1982) Influence of social-category activation on processing of visual information. *Social Cognition, 1:2,* 95-109.

Kreiman, Jody. (1997) Listening to voices: Theory and practice in voice perception research. In Keith Johnson & John W. Mullennix (Eds.), *Talker Variability in Speech Processing* (pp. 85-108). San Diego, California: Academic Press.

Kruskal, Joseph B. & Myron Wish. (1978) *Multidimensional Scaling.* Newbury Park: Sage Publications.

Kuhl, Patricia K. & Andrew N. Meltzoff. (1984) The intermodal representation of speech in infants. *Infant Behavior & Development, 7,* 361-381.

Kuhl, Patricia K. & Andrew N. Meltzoff. (1982) The bimodal perception of speech in infancy. *Science, 218,* 1138-1141.

Labov, William. (1994) *Principles of Linguistic Change.* Cambridge, Massachusetts: Blackwell Publishers.

Ladefoged, Peter & D.E. Broadbent. (1957) Information conveyed by vowels. *Journal of the Acoustical Society of America, 29:1,* 98-104.

Lambert, Wallace. (1967) A social psychology of bilingualism. *Journal of Social Issues, 23:2,* 91-109.

Lambert, Wallace, R. Hodgson, R. Gardner, & Samuel Fillenbaum. (1960) Evaluative reactions to spoken language. *Journal of Abnormal and Social Psychology, 60,* 44-51.

Langlois, Judith H. & Lori A. Roggman. (1990) Attractive faces are only average. *Psychological Science, 1,* 115-121.

Laqueur, Thomas. (1990) *Making Sex: Body and Gender from the Greeks to Freud.* Cambridge, Massachusetts: Harvard University Press.

Lemm, Kristi M., M. Dabady, Mahzarin R. Banaji, & J. Kluewer. (1999) The influence of gendered language on thought. Paper presented at the annual meeting of the Midwestern Psychological Association, Chicago, Illinois.

MacDonald, John W. & Harry McGurk. (1978) Visual influences on speech perception processes. *Perception & Psychophysics, 24,* 253-257.

Manis, Melvin, Thomas E. Nelson, & Jonathan Shedler. (1988) Stereotypes and social judgment: Extremity, assimilation, and contrast. *Journal of Personality and Social Psychology, 55:1,* 28-36.

Markus, Hazel. (1977) Self-schemata and processing information about the self. *Journal of Personality and Social Psychology, 35,* 63-78.

Martin, M. Kay & Barbara Voorhies. (1975) *Female of the Species.* New York: Columbia University Press.

McArthur, Leslie Z. & Karen Apatow. (1983/4) Impressions of baby-faced adults. *Social Cognition, 2,* 315-342.

McConahay, John B. & Joseph C. Hough, Jr. (1976) Symbolic racism. *Journal of Social Issues, 32:2,* 23-45.

McGurk, Harry & John MacDonald. (1976) Hearing lips and seeing voices. *Nature, 264,* 746–48.

McKenzie-Mohr, Doug & Mark P. Zanna. (1990) Treating women as sexual objects: Look to the (gender schematic) male who has viewed pornography. *Personality and Social Psychology Bulletin, 16,* 296-308.

Medin, Douglas L. & Marguerite M. Schaffner. (1978) Context theory of classification learning. *Psychological Review, 85:3,* 207-238.

Miller, Cynthia L., Barbara A. Younger, & Philip A. Morse. (1982) The categorization of male and female voices in infancy. *Infant Behavior and Development, 5,* 143-149.

Miller, James D. (1989) Auditory-perceptual interpretation of the vowel. *Journal of the Acoustical Society of America, 85,* 2114-2134.

Mills, Carol J. (1983) Sex-typing and self schemata effects on memory and response latency. *Journal of Personality & Social Psychology, 45:1,* 163-172.

Mitford, Jessica. (1993) *The American Way of Birth.* New York: Plume.

Mullennix, John W., Keith A. Johnson, Meral Topcu-Durgun, & Lynn M. Farnsworth. (1995) The perceptual representation of voice gender. *Journal of the Acoustical Society of America, 98:6,* 3080-3095.

Mullennix, John W., David B. Pisoni, & Christopher S. Martin. (1988) Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America, 85:1,* 365-378.

Munhall, Kevin G., P. Gribble, L. Sacco, & M. Ward. (1996) Temporal constraints on the McGurk effect. *Perception & Pscyhophysics, 58:3,* 351-362.

Murry, Thomas & Sadanand Singh. (1980) Multidimensional analysis of male and female voices. *Journal of the Acoustical Society of America, 68:5,* 1294-1300.

Neely, James H. (1977) Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General, 106,* 226-254.

Nicholson, Linda. (1994) Interpreting "gender." *Signs: Journal of Women in Culture and Society, 20:1,* 79-105.

Niedzielski, Nancy. (1999) The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology, 18:1,* 62–85.

Niedzielski, Nancy. (1997) *The Effect of Social Information on the phonetic perception of sociolinguistic variables.* Ph.D. Dissertation, University of California. Santa Barbara, California.

Nolan, Francis. (1995) The Handbook of the International Phonetic Association. *Journal of the International Phonetic Association, 25:1.*

Nusbaum, Howard C., David B. Pisoni, & Christopher K. Davis. (1984) Sizing up the Hoosier Mental Lexicon: Measuring the familiarity of 20,000 words. *Research on Speech Perception Progress Report No. 10, Indiana University Psychology Department*, 357-376.

Nygaard, Lynne C. & David B. Pisoni. (1998) Talker-specific learning in speech perception. *Perception & Psychophysics, 60:3,* 355-376.

Nygaard, Lynne C., Mitchell S. Sommers, & David B. Pisoni. (1994) Speech perception as a talker-contingent process. *Psychological Science, 5:1,* 42-46.

Pisoni, David B. (1997) Some thoughts on "normalization" in speech perception. In Keith Johnson & John W. Mullennix (Eds.), *Talker Variability in Speech Processing* (pp. 9-32). San Diego, California: Academic Press.

Posner, Michael I. & Charles R.R. Snyder. (1975) Facilitation and inhibition in the processing of signals. In Patrick Rabbitt & Stanislav Dornic (Eds.), *Attention and Performance* (Volume 5; pp. 669-682). New York: Academic Press.

Preston, Dennis R. (1999) *Handbook of Perceptual Dialectology* (Volume 1). Philadelphia: John Benjamins Publishing.

Purnell, Thomas, William Idsardi, & John Baugh. (1999) Perceptual and phonetic experiments on American English dialect identification. *Journal of Language and Social Psychology, 18:1,* 10-30.

Rosch, Eleanor. (1978) Principles of categorization. In Eleanor Rosch & B.B. Lloyd (Eds.), *Cognition and Categorization* (pp. 27-48). Hillsdale, NJ: Erlbaum.

Sekiyama, Kaoru. (1997) Cultural and linguistic factors in audiovisual speech processing: the McGurk effect in Chinese subjects. *Perception & Psychophysics, 59:1,* 73-80.

Sekiyama, Kaoru. (1994) Differences in auditory-visual speech perception between Japanese and Americans: McGurk effect as a function of incompatibility. *Journal of the Acoustical Society of Japan, 15,* 143-148.

Sekiyama, Kaoru, Louis D. Braida, Keiko Nishino, Mieko Hayashi, & Michael M. Tuyo. (1995) The McGurk effect in Japanese and American perceivers. *Proceedings of the International Congress of Phonetic Sciences, Stockholm, 4,* 214-217.

Sekiyama, Kaoru & Yoh'ichi Tohkura. (1993) Inter-language differences in the influence of visual cues in speech perception. *Journal of Phonetics, 21,* 427-444.

Sekiyama, Kaoru & Yoh'ichi Tohkura. (1991) McGurk effect in non-English listeners: Few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility. *Journal of the Acoustical Society of America, 90:4, Part 1,* 1797-1805.

Smith, Eliot R. (1998) Mental representation and memory. In Daniel T. Gilbert, Susan T. Fiske, & Gardner Lindzey (Eds.), *The Handbook of Social Psychology* (Volume 1; pp. 391-444). New York: Oxford University Press.

Smith, Eliot R. & M.A. Zarate. (1992) Exemplar-based model of social judgment. *Psychological Review, 99,* 3-21.

Snyder, Mark & Peter Miene. (1994) On the functions of stereotypes and prejudice. In Mark P. Zanna & James M. Olson (Eds.), *The Psychology of Prejudice: The Ontario Symposium* (Volume 7; pp. 33-54). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Sowell, Thomas. (1994) *Race and Culture: A World View.* New York: BasicBooks.

Sowell, Thomas. (1981) *Ethnic America: A History.* New York: BasicBooks.

Spence, Janet T., Robert L. Helmreich, & Jay Stapp. (1975) Ratings of self and peers on sex role attributes and their relation to self-esteem and conceptions of masculinity and femininity. *Journal of Personality and Social Psychology, 32,* 29-39.

Stangor, Charles, Laure Lynch, Changming Duan, & Beth Glas. (1992) Categorization of individuals on the basis of multiple social features. *Journal of Personality & Social Psychology, 62:2,* 207-218.

Strand, Elizabeth A. (1999) Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Social Psychology, 18:1,* 86-100.

Strand, Elizabeth A. & Keith Johnson. (1996) Gradient and visual speaker normalization in the perception of fricatives. In Dafydd Gibbon (Ed.), *Natural Language Processing and Speech Technology: Results of the 3rd KONVENS Conference, Bielefeld, October 1996* (pp. 14-26). Berlin: Mouton de Gruyter.

Summerfield, Quentin. (1987) Some preliminaries to a comprehensive account of audio-visual speech perception. In Barbara Dodd & Ruth Campbell (Eds.), *Hearing by Eye: The Psychology of Lip Reading* (pp. 3-51). Hove, UK: Lawrence Erlbaum Associates Ltd.

Swim, Janet K., Kathryn J. Aikin, Wayne S. Hall, & Barbara A. Hunter. (1995) Sexism and racism: Old-fashioned and modern prejudices. *Journal of Personality and Social Psychology, 68:2,* 199-214.

Syrdal, Anne & H. Gopal. (1986) A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America, 79,* 1086-1100.

Taylor, Shelley E. (1981) A categorizational approach to stereotyping. In David L. Hamilton (Ed.), *Cognitive Processes in Stereotyping and Intergroup Behavior* (pp. 83-114). Hillsdale, New Jersey: Erlbaum.

Taylor, Shelley E., Letitia Anne Peplau, & David O. Sears. (1994) *Social Psychology.* Englewood Cliffs, New Jersey: Prentice Hall.

Terman, Lewis M. & Catherine Cox Miles. (1936) *Sex and Personality: Studies in Masculinity and Femininity.* New York: McGraw-Hill.

Thakerar, Jitendra N. & Howard Giles. (1981) They are—so they spoke: Noncontent speech stereotypes. *Language & Communication, 1:2/3,* 255-261.

Tougas, Francine, Rupert Brown, Ann M. Beaton, & Stephane Joly. (1995) Neosexism: Plus ça change, plus c'est pareil. *Personality and Social Psychology Bulletin, 21:8,* 842-849.

Traunmüller, Hartmut. (1994) Conventional, biological, and environmental factors in speech communication: A modulation theory. *Phonetica, 51,* 170-183.

Traunmüller, Hartmut. (1981) Perceptual dimension of openness in vowels. *Journal of the Acoustical Society of America, 69,* 1465-1475.

Trzebinski, Jerzy & Kathleen Richards. (1986) The role of goal categories in person impression. *Journal of Experimental Social Psychology, 22:3,* 216-227.

Trzebinski, Jerzy. (1985) Action-oriented representations of implicit personality theories. *Journal of Personality and Social Psychology, 48:5,* 1266-1278.

Unger, Rhoda K. & Mary Crawford. (1993) Commentary: Sex and Gender—The troubled relationship between terms and concepts. *Psychological Science, 4:2,* 122-124.

Valentine, Tim. (1991a) A unified account of the effects of distinctiveness, inversion, and race on face recognition. *Quarterly Journal of Experimental Psychology, 43A:2,* 161-204.

Valentine, Tim. (1991b) Representation and process in face recognition. In R. Watt (Ed.), *Vision and Visual Dysfunction: Volume 14. Pattern Recognition in Man and Machine.* London: Macmillan.

Vaughter, Reesa M. (1976) Review essay: Psychology. *Signs, 2,* 120-146.

Walker-Andrews, Arlene S., Lorraine E. Bahrick, Stacy S. Raglioni, & Isabel Diaz. (1991) Infants' bimodal perception of gender. *Ecological Psychology, 3:2,* 55-75.

Wardaugh, Ronald. (1992) *An Introduction to Sociolinguistics* (Second Edition). Oxford: Blackwell Publishers.

Weisstein, Naomi. (1968) *Kinder, Kuche, Kirche as Scientific Law: Psychology Constructs the Female*. Boston: New England Free Press.

Wilkinson, Leland, MaryAnn Hill, & Erin Vang. (1992) *SYSTAT: Statistics, Version 5.2 Edition*. Evanston, Illinios: SYSTAT, Inc.

Wish, Myron. (1970) Comparisons among multidimensional structures of nations based on different measures of subjective similarity. *General Systems, 15,* 55-65.

Wish, Myron & J. Douglas Carroll. (1974) Applications of individual differences scaling to studies of human perception and judgment. In Edward C. Carterette & Morton P. Friedman (Eds.), *Handbook of Perception* (Volume 2; pp. 449-491). New York: Academic Press.

Wolf, Naomi. (1991) *The Beauty Myth: How Images of Beauty Are Used Against Women*. New York: Doubleday.

Yamaguchi, Masami K., Tastu Hirukawa, & So Kanazawa. (1995) Judgment of gender through facial parts. *Perception, 24,* 563-575.

Zarate, Michael A. & Eliot R. Smith. (1990) Person categorization and stereotyping. *Social Cognition, 8:2,* 161-185.

Zebrowitz, Leslie A. (1997) *Reading Faces: Window to the Soul?* Boulder, Colorado: Westview Press.

Zebrowitz-McArthur, Leslie & Diane S. Berry. (1987) Cross-cultural agreement in perceptions of babyfaced adults. *Journal of Cross-Cultural Psychology, 18:2,* 165-192.

Zebrowitz-McArthur, Leslie & Joann M. Montepare. (1989) Contributions of a babyface and a childlike voice to impressions of moving and talking faces. *Journal of Nonverbal Behavior, 13:3,* 189-203.