TILBURG ◆ UNIVERSITY

Tilburg University

Stereotyping based on voice in the presence of individuating information

Ko, S.J.; Judd, C.M.; Stapel, D.A.

Published in:
Personality and Social Psychology Bulletin

Publication date:
2009

Link to publication in Tilburg University Research Portal

Citation for published version (APA):
Ko, S. J., Judd, C. M., & Stapel, D. A. (2009). *Stereotyping based on voice in the presence of individuating information: Vocal femininity affects perceived competence but not warmth.* Manuscript in preparation.

Download date: 12. May. 2021

Personality and Social Psychology Bulletin

Stereotyping Based on Voice in the Presence of Individuating Information: Vocal Femininity Affects Perceived Competence but Not Warmth

Sei Jin Ko, Charles M. Judd and Diederik A. Stapel

Published by:

$SAGE

http://www.sagepublications.com

On behalf of:

SP
SP

Society for Personality and Social Psychology, Inc.

Additional services and information for *Personality and Social Psychology Bulletin* can be found at:

Email Alerts: http://psp.sagepub.com/cgi/alerts

Subscriptions: http://psp.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations http://psp.sagepub.com/cgi/content/refs/35/2/198

Stereotyping Based on Voice in the Presence of Individuating Information: Vocal Femininity Affects Perceived Competence but Not Warmth

Sei Jin Ko
Northwestern University

Charles M. Judd
University of Colorado at Boulder

Diederik A. Stapel
Tilburg University, the Netherlands

In two experiments the authors examined the effect of vocal cues on warmth and competence judgments when other competing information was concurrently available. In Experiment 1, using male and female speakers posing as job applicants, the authors investigated how applicants' vocal cues and résumé information impacted judgments of competence and warmth. Results showed competence was solely affected by vocal femininity—applicants with masculine voices were rated as more competent than applicants with feminine voices, regardless of applicant gender or résumé information. Warmth was predominantly affected by résumés—applicants with feminine résumés were rated as warmer than applicants with masculine résumés. In Experiment 2, the potent effect of vocal femininity on competence was replicated even under conditions where the competing background information was directly diagnostic of warmth and competence. Furthermore, the authors found that the impact of vocal femininity on competence was largely due to the overlap between perceptions of vocal femininity and babyishness.

Keywords: voice; vocal femininity; vocal babyishness; individuating information; gender; stereotyping

In recent years a new line of work has emerged that provides a richer account of the underlying basis for stereotyping. Specifically, recent empirical evidence suggests that stereotypes are based not only on inferences about the category membership of stereotyped individuals (e.g., African American vs. White individuals) but also on physical cues or features that are associated with category membership but that concurrently vary within each category (e.g., Blair, Judd, & Chapleau, 2004; Blair, Judd, & Fallman, 2004; Blair, Judd, Sadler, & Jenkins, 2002; Eberhardt, Davies, Purdie-Vaughns, & Johnson, 2006; Livingston & Brewer, 2002; Maddox & Gray, 2002). So for instance, both Whites and African Americans who are clearly identified as members of one category or the other, but who vary in Afrocentric facial

features, also vary in the stereotypic attributes that they are seen to possess. Within each category, those with more Afrocentric features are judged to possess attributes that are more stereotypic of African Americans and less stereotypic of Whites.

It is clear from this past work that physical features can and do affect stereotypic perceptions. Nonetheless, life is not like a silent movie in which we are only provided with images of a target individual. Indeed, our social world is made up of more than what we soak in through our eyes. Undoubtedly, the visual aspects of the actors in silent movies—visual cues that are diagnostic of race and gender, for instance—provide a rich and fascinating basis for stereotypic inferences, but what if the actors' silent mouth movements gave way to audible vocalizations? Would their vocal cues also lead perceivers to stereotype them? Just thinking about how vividly and potently voices on the radio or the telephone can conjure up images of the speakers suggests that the anecdotal answer to this question is a resounding yes.

The empirical demonstration of this is only just starting to emerge. Ko, Judd, and Blair (2006) recently provided the first demonstration of stereotyping on the basis of vocal cues that varied both between categories and within them. Specifically, these researchers showed that solely as a function of listening to brief vocal clips, participants had little trouble making gender-stereotypic inferences based not only on the speaker's gender (as revealed by the voice) but also on variations in vocal qualities *within* gender, which they refer to as "vocal femininity." In other words, speakers with more feminine sounding voices were seen as more stereotypically feminine than speakers with less feminine sounding voices *regardless* of the speaker's gender.

Further, Ko, Muller, Judd, and Stapel (2008) examined the relationship between category-based and feature-based stereotyping in order to illustrate how Western societal norms encouraging the suppression of category-based stereotyping can ironically exacerbate biases through a rebound in within-category feature-based stereotyping. Specifically, participants in their study engaged in a standard suppression writing task (see Macrae, Bodenhausen, Milne, & Jetten, 1994) where they wrote about a male and a female target following either instructions to avoid the use of gender category–based stereotypes (suppression condition) or no such instructions (control condition). Stereotyping was measured via a postsuppression voice judgment task. They found that participants in the suppression condition, as compared to those in the control, continued to curb their category-based stereotyping (category was derived from voice) while simultaneously relying more heavily on vocal femininity cues (within gender category) to make stereotypic inferences. A particularly interesting aspect of this research was its demonstration that suppression

in one modality (participants were told to suppress their use of gender stereotypes based on visually presented stimulus person) can affect the degree to which perceivers used feature-based stereotyping in another modality (they engaged in greater feature-based stereotyping based on auditory cues).

Even though these were some of the first demonstrations of within-category stereotyping on the basis of voice—particularly on the basis of vocal femininity—one could still argue that because perceivers were only presented with vocal cues in these studies, it was only logical that they used these cues in their inferences. Thus, a more compelling case for the role of voice in stereotyping would be the demonstration of its impact alongside other competing information. Accordingly, the purpose of the present research is to examine whether vocal cues continue to exert an influence on stereotypic judgments even when other rich competing cues are concurrently available.

In our examination we wanted to go further than simply testing whether vocal cues continue to impact stereotypic judgments in the presence of other information. We more specifically wanted to know whether there are particular dimensions of judgment that may be uniquely affected by the voice when perceivers are presented with both the speakers' voice and individuating information cues. In order to answer this question, one needs to consider first the distinctions that fundamentally differentiate the content of stereotypic judgments.

FUNDAMENTAL DIMENSIONS OF SOCIAL JUDGMENT: WARMTH AND COMPETENCE

It appears that there is fairly consistent agreement among researchers that there are two fundamental dimensions that underlie social judgment, regardless of whether one is judging individuals or social groups (Bakan, 1966; Fiske, Cuddy, Glick, & Xu, 2002; Fiske, Xu, Cuddy, & Glick, 1999; Heilman, 2001; Judd, James-Hawkins, Yzerbyt, & Kashima, 2005; Rosenberg, Nelson, & Vivekananthan, 1968). Though different researchers have attached slightly different labels to these dimensions, we will adopt the labels widely used by Fiske and colleagues (1999, 2002) in their stereotype content model: *warmth* and *competence*. Warmth is predominantly associated with other or service orientation and hence represented by positive stereotypic female versus negative stereotypic male traits (e.g., caring and friendly vs. cold and uncaring). Competence, on the other hand, is predominantly associated with achievement orientation and hence represented by positive stereotypic male versus negative stereotypic female traits (e.g., intelligent and motivated vs. lazy and incompetent; Cejka & Eagly, 1999; Eagly & Karau, 2002;

Eagly & Steffen, 1984; Fiske et al., 2002). These two dimensions have been particularly effective at capturing the content of people's gender-stereotypic beliefs. This is because the content of gender stereotypes associated with each gender group is not uniformly negative or positive. Rather, positive perceptions or high ratings on one dimension are juxtaposed with negative perceptions or low ratings on the other dimension. Which dimension yields high and which low ratings depends on the status of the group in question. That is, men, the higher status group, are rated high on competence but low on warmth, whereas women, the lower status group, are rated low on competence but high on warmth (Fiske et al., 2002). Consequently, these dimensions are able to explain the seemingly paradoxical coexistence of positive stereotypes about women yet persistent gender prejudice (Eagly & Mladinic, 1989): Women are described with positive warmth-related virtues yet discriminated against because of attributed incompetence.

The question that we turn to is whether vocal cues, in the presence of other target information, affect both of these dimensions equally or one more strongly than the other. Under situations where vocal cues are the sole information about a target person, it makes sense that vocal cues would impact both dimensions of judgment, since those are the only cues available and since gender stereotypes vary along both dimensions. However, in situations where vocal cues are just one among other concurrently available sources of target information, it is unclear how vocal cues may impact judgments. Perhaps the most logical prediction is that vocal cues and competing individuating information would affect both dimensions equally, insofar as both sources of target information are gender stereotypic. The specific nature of these effects would be that applicants with more masculine voices and/or masculine résumés may be rated as more competent but less warm than those with more feminine voices and/or feminine résumés. However, there is suggestive evidence from a variety of literatures that points to an alternative possibility. Specifically, this evidence suggests that vocal cues may be perceived to be more diagnostic of and have a stronger impact on competence than warmth. In the next section we detail the rationale for this alternative prediction and review the relevant research.

IMPACT OF VOICE ON COMPETENCE

Years of gender research have taught us that perceived gender is a multifaceted construct that is shaped by a myriad of different factors. Specifically, perceptions of femininity/masculinity are influenced by a myriad of factors such as traits, role behaviors, occupations, and physical attributes such as facial and vocal features (e.g., Deaux & Lewis, 1984; Eagly, 1987; Eagly & Steffen, 1984; Glick, Wilk, & Perreault, 1995; Glick, Zion, & Nelson, 1988; Ko et al., 2006, 2008; Spence, 1993). These factors represent different facets of femininity/masculinity. The facet of femininity/masculinity represented by a person's physical features or cues may have considerable overlap with the person's babyishness. Indeed, Zebrowitz and colleagues' research on "babyfacedness" (Berry & McArthur, 1986; Friedman & Zebrowitz, 1992; Montepare & Zebrowitz, 1998) has corroborated this by showing that many of the physical features that distinguish babies from adults (e.g., big eyes, hairless face, small jaw) are also those that distinguish adult women from adult men. To the extent that this is true also for vocal characteristics, vocal femininity cues will be more strongly associated with competence than warmth since babies and adults necessarily are seen to differ primarily on competence rather than on warmth. Support for this possibility comes from work by Zebrowitz and Montepare (2005) who found a high negative correlation between babyfacedness and competence in recent congressional and senatorial elections. In contrast, differing roles adopted by men and women may represent other facets of femininity/masculinity. In particular, social role theory (Eagly, 1987) argues that the stereotypic association between women and (high) warmth results from the preponderance of women in caregiver roles (e.g., nurse, elementary school teacher, babysitter). Accordingly, perceivers may see information about target's role behavior (e.g., past jobs, hobbies) as more diagnostic of warmth than competence.

Although no research to date has directly examined how the voice impacts the fundamental dimensions of social perceptions—especially in the face of other target information—a handful of disparate research provides the impetus for the alternative reasoning involving vocal cues and competence. For instance, personality research has shown that people who speak more quickly and have more variability in pitch are seen as more competent (Brown, Strong, & Rencher, 1973; Smith, Brown, Strong, & Rencher, 1975; also see Aronovitch, 1976) and have more persuasive influence (Miller, Maruyama, Beaber, & Valone, 1976). In the area of communication studies, Tusing and Dillard (2000) demonstrated a relationship between voice acoustics and dominance perceptions under conditions where other cues were concurrently available. Specifically, these researchers found that vocal cues of loudness and variability in loudness were positively correlated and pitch was negatively correlated with dominance judgments even when controlling for the effects derived from speech content and visual cues of target speakers. Perhaps even more

germane to our purpose is work by DeGroot and Motowidlo (1999) who examined the relationship between ratings of targets' job and interview performance and their visual and vocal cues. For this purpose, the researchers video- and audiotaped interviews of male and female target employees from a large news publishing company (Study 2). The targets were then rated on job and interview performance and on a number of visual and vocal characteristics. They found that targets' vocal cues were more prominently associated with how they were judged on job and interview performance than were their visual cues.

In sum, there is literature showing that vocal cues are influential in perceptions of dominance, competence, persuasion, and job performance, even in the face of other competing information or cues. Given that these are all traits that fall under the competence dimension, the findings clearly hint at a potentially prominent role of vocal cues in competence judgments. However, it is not possible to draw more definitive conclusions from this past research because the designs were not set up to make systematic comparisons between competence and warmth judgments based on voice. In other words, because the dependent measures were predominantly restricted to items measuring the competence dimension, one could argue that the findings were, in part, an artifact of the research design. Consequently, whether vocal cues impact warmth is an open question. Furthermore, this past work does not directly inform us about gender stereotyping, whereas the current work does by focusing on how perceptions of vocal cues associated with gender—vocal femininity—impacts the content of stereotypic judgments.

EXPERIMENT 1: OVERVIEW

In Experiment 1, we examined how the interplay of targets' vocal and background information cues affect judgments of them on warmth and competence. For this purpose we modified a research paradigm used by Glick et al. (1988). Specifically, applicants' background information was presented through résumé-type information, which depicted either a stereotypically masculine or a stereotypically feminine applicant. For the voices, we recorded male and female speakers posing as applicants and had their voices scaled on vocal femininity/masculinity. On the basis of this scaling, male and female applicants whose voices were rated as either high or low on femininity (for their gender) were chosen for the main part of the experiment, where vocal femininity was crossed with voice gender category and résumé sex type such that each participant heard a different voice-résumé pairing. This allowed us to assess the nature of the separate

and joint effects of vocal femininity, résumés, and speaker gender on the judgments made of the speakers. The dependent variables were judgments of each speaker's levels of competence and warmth.

Like past research on individuating/résumé information (e.g., Beckett & Park, 1995; Glick et al., 1988; Locksley, Borgida, Brekke, & Hepburn, 1980; Locksley, Hepburn, & Ortiz, 1982; also see Davison & Burke, 2000), we expected résumé information to affect judgments of speakers. More importantly for our purposes, we anticipated that vocal femininity would influence judgments (Ko et al., 2006), while gender category itself may not. Our rationale for these predictions comes from recent work suggesting that people in Western societies may have become quite good at controlling the more blatant category-based stereotyping but not the more subtle within-category feature-based stereotyping, particularly in situations where people may be wary of appearing biased (Blair, Judd, & Chapleau, 2004; Eberhardt et al., 2006; Ko et al., 2008). Intuitively, the most plausible prediction was that résumés and vocal cues would impact the two dimensions of judgments equally. However, given the suggestive evidence reviewed previously, it seems possible that vocal cues would have a more powerful influence on competence than warmth, in which case résumés may impact warmth more strongly than competence.

EXPERIMENT 1: EXPERIMENTAL MATERIALS

Résumés: Construction and Selection

We constructed 24 sex-typed résumés that were designed to depict either masculine or feminine applicants. The résumés' sex-type was manipulated by varying three pieces of information that made up each résumé. These three pieces were (a) a holiday job (e.g., masculine job: salesperson at an auto supply store; feminine job: sales assistant at a flower shop), (b) a job held during university term time (e.g., masculine job: security patrol staff of a company; feminine job: an aerobics instructor), and (c) a favorite hobby (e.g., masculine hobby: playing basketball; feminine hobby: yoga). The three pieces of information contained in each résumé were always stereotypically consistent with each other. In other words, if a résumé started with a masculine holiday job, then the term time job and the favorite hobby were also masculine.[1]

These résumés were pretested for stereotypicality with 19 participants from the University of Groningen who rated them on a scale ranging from 1 (*very feminine*) to 7 (*very masculine*). Consequently, we chose six feminine sex-typed résumés ($M = 1.79$) and six masculine sex-typed résumés ($M = 6.25$).[2]

Applicant Voices: Recording and Voice Scaling

To construct our materials, we recorded male and female participants reading the résumés as if they were the actual applicants depicted in the résumés. Each of these participants also read a neutral text called the *Rainbow Passage* (Fairbanks, 1940, p. 127; see also Ko et al., 2006). This passage was used to scale the voices on femininity and the resulting scale values were used to select the voices used in the experiment.

In all, 85 students (45 female and 40 male) at the University of Groningen were recorded in exchange for monetary compensation. All were native Dutch speakers with an average age of 23 years ($SD = 2.40$). We recorded the participants in a soundproofed recording room. During the recording session, each participant was asked to read aloud the *Rainbow Passage* and each of the 12 résumés. For each résumé, male participants always began by stating a different male name (e.g., "My name is Mark") and female participants always began by stating a different female name (e.g., "My name is Ellen"). We had to discard 6 of the female and 2 of the male recordings because they contained too much background noise or too many reading errors. On the basis of the *Rainbow Passage*, the remaining 39 female and 38 male voice recordings were then scaled on vocal femininity.

In all, 79 undergraduate students (72 female and 7 male) at the University of Groningen provided judgments of the vocal femininity of the recorded voices in exchange for partial fulfillment of course credit. Participants were seated in front of individual computers with headphones and assigned to judge either the 38 male or 39 female voices speaking the *Rainbow Passage*. The order of the voices varied randomly across participants. The participants' task was to rate each voice on how feminine it sounded, using a 1 (*not at all feminine*) to 7 (*very feminine*) scale.[3]

Voice scaling results. To examine the reliabilities of the mean femininity scores for each voice, the individual femininity ratings (for each voice by each participant) were analyzed with a two-way analysis of variance within the male and female voice sets, treating judges/participants as one factor and voice as another. From these analyses, variance components were estimated and intraclass correlations (ICC; due to voice, treating participants as a random factor; ICC 2.1 from Shrout & Fleiss, 1979) were calculated in order to estimate the reliabilities of the resulting vocal femininity values (following procedures outlined in Judd & McClelland, 1998). For the female voices, the mean femininity score was 4.56 ($SD = .55$) and the reliability of the individual voice means was .92. For the male voices the mean femininity score was 3.10

($SD = .63$) and the reliability of the individual voice means was .97. As one would expect, the female voices were rated as significantly more feminine than the male voices, $F(1, 77) = 120.25$, $p < .001$.

On the basis of the vocal femininity judgments 3 female applicants relatively high on vocal femininity ($M=5.56$), 3 female applicants relatively low on vocal femininity ($M=3.33$), 3 male applicants relatively high on vocal femininity ($M=4.90$), and 3 male applicants relatively low on vocal femininity ($M=1.90$) were chosen to represent the final 12 applicant speakers for the main experiment.

EXPERIMENT 1: RATINGS OF APPLICANTS

Our goal was to examine how the applicants would be judged on the dimensions of warmth and competence. Accordingly, participants were presented with a series of hypothetical applicants, some in written form only (with their résumé and gender provided) and some orally (varying in résumé type, gender, and vocal femininity [within gender]). Each applicant was then judged by each participant on rating scales relevant to the competence and warmth dimensions.

Method

Participants. In all, 62 undergraduate students (40 female and 22 male) at the University of Groningen participated for partial fulfillment of course credit or monetary compensation.

Research design. Based on pretesting, as described earlier, we selected a total of 12 résumés—6 stereotypically masculine and 6 stereotypically feminine. Participants in this experiment evaluated the applicants described by these résumés. Of the résumés, 4 were judged in text form only, with participants making their judgments on the basis of reading the résumés. The text-only presentation served as a control to gauge the effect of written information without any influence coming from the physical cue of voice. Of the résumés, 8 were associated with voices, such that each résumé was orally presented by a different speaker, the speaker ostensibly self-presenting himself or herself. In the oral presentation, each participant was presented with two randomly selected voices out of the total three per voice type (e.g., male applicants low on femininity). However, across participants, all three voices per voice type were presented. Which résumés were text only and which were orally presented was counterbalanced across participants.

The four text-only résumés varied according to a 2 (applicant gender category: female vs. male) × 2 (résumé

sex-type: feminine vs. masculine) within-participant design. The first factor was accomplished by varying the name at the top of the résumé. A further counterbalancing here was done such that each of the two stereotypically feminine résumés, for instance, was attributed to a male applicant for half the participants and to a female applicant for the other half of the participants.

For the eight résumés that were orally presented, the design involved 2 (speaker gender: female vs. male) × 2 (résumé sex-type: feminine vs. masculine) × 2 (vocal femininity: high vs. low within gender) within-participant factors. Here we also did extensive counterbalancing across participants so that each résumé was equally often presented by a male or a female and equally often by a high feminine voice (within gender) or a low feminine voice (again within gender).

Dependent measures of warmth and competence. Each applicant, either presented in text only or orally, was rated by participants on eight traits. Four of these focused on the competence dimension, two at the positive end (assertive and decisive) and two at the negative end (passive and insecure). The other four referred to the warmth dimension, again with two positive warmth traits (supportive and caring) and two negative warmth traits (cold and hostile). Each trait rating was recorded using a 7-point scale (1 = *not at all* [trait] to 7 = *very* [trait]). From these eight ratings of each stimulus résumé, we derived two scores, one on perceived competence and one on perceived warmth, in both cases subtracting the two ratings on the traits at the negative end of the dimension from the two ratings on the traits at the positive end.[4]

Procedure. On arrival, participants were seated in front of individual computers with headphones. The computer automatically assigned each participant to the appropriate counterbalancing condition. All instructions were presented via the computer. The instructions informed the participants that the experiment was about examining how closely the lay public's impressions of applicants (as represented by their impressions) matched impressions made by experienced job recruiters. As a rationale for the text presentation of the résumés (as opposed to the oral presentation), participants were given the cover story that a few of the audiotapes of applicant résumés turned out to be corrupt and consequently, only the transcripts of the résumés would be presented for these applicants.

After the instructions, the computer randomly presented the 12 résumés one after the other, either spoken by applicant voices or in text format. After each résumé, participants rated the applicant on the eight traits. The order of these was randomized for each participant.

While the ratings were being done for each applicant, either the spoken résumé looped or the text résumé stayed on the screen.

Results

Judgments of résumés presented in text only. In order to examine how applicants' gender category (derived from name) and/or résumé information affected judgments, the composite ratings of résumés presented in text form were subjected to a 2 (applicant gender: female vs. male) × 2 (judgment dimension: warmth vs. competence) × 2 (résumé sex-type: feminine vs. masculine) within-participants ANOVA.[5] Presented in Table 1 are the two dimensions' means and standard deviations, broken down by gender category and résumé type. There was a category main effect, $F(1, 61) = 4.97$, $p < .05$, due to higher average ratings given to female than male applicants; a judgment dimension main effect, $F(1, 61) = 4.84$, $p < .05$, due to higher average ratings on warmth than on competence; and a résumé sex-type main effect, $F(1, 61) = 35.53$, $p < .0001$, due to higher average ratings for feminine than masculine résumés. The only significant interaction was between judgment dimension and résumé, $F(1, 61) = 138.40$, $p < .0001$, showing that feminine résumés were rated higher on warmth while masculine ones were rated higher on competence. In sum, female applicants were rated more highly than males on both dimensions, while résumé sex type seemed to be matched with judgmental dimension (i.e., masculine résumés were rated higher on competence but lower on warmth than feminine résumés). The absence of a gender category by judgment dimension interaction suggests that these judgments were not based on stereotyping by gender exclusively.

Judgments of résumés presented orally. The composite ratings for the oral presentation were subjected to a 2 (applicant gender: female vs. male) × 2 (applicant vocal femininity: high feminine vs. low feminine) × 2 (judgment dimension: warmth vs. competence) × 2 (résumé sex-type: feminine vs. masculine) within-participants ANOVA.

Presented in Figure 1 are the average ratings on the two dimensions of competence and warmth as a function of résumé type, gender category, and vocal femininity. There were main effects of gender category, $F(1, 61) = 82.69$, $p < .0001$, due to higher ratings for female than male applicants; vocal femininity, $F(1, 61) = 33.09$, $p < .0001$, due to higher ratings for low than high feminine voices; judgment dimension, $F(1, 61) = 30.71$, $p < .0001$, due to higher ratings on warmth than competence; and résumé, $F(1, 61) = 38.92$, $p < .0001$, due to higher ratings for feminine than masculine résumés.

TABLE 1: Experiment 1: Text Presentation of Résumés—Means (and Standard Deviations) for the Composite Ratings on the Dimensions of Warmth and Competence, Broken Down by Applicant Gender and Résumé Type

	Dimensions							
	Competence				Warmth			
	Female Applicant		Male Applicant		Female Applicant		Male Applicant	
Résumé Type	M	SD	M	SD	M	SD	M	SD
Feminine	4.87	.95	4.54	.94	5.63	.77	5.68	.62
Masculine	5.10	.78	4.93	.88	4.44	.78	4.31	.93
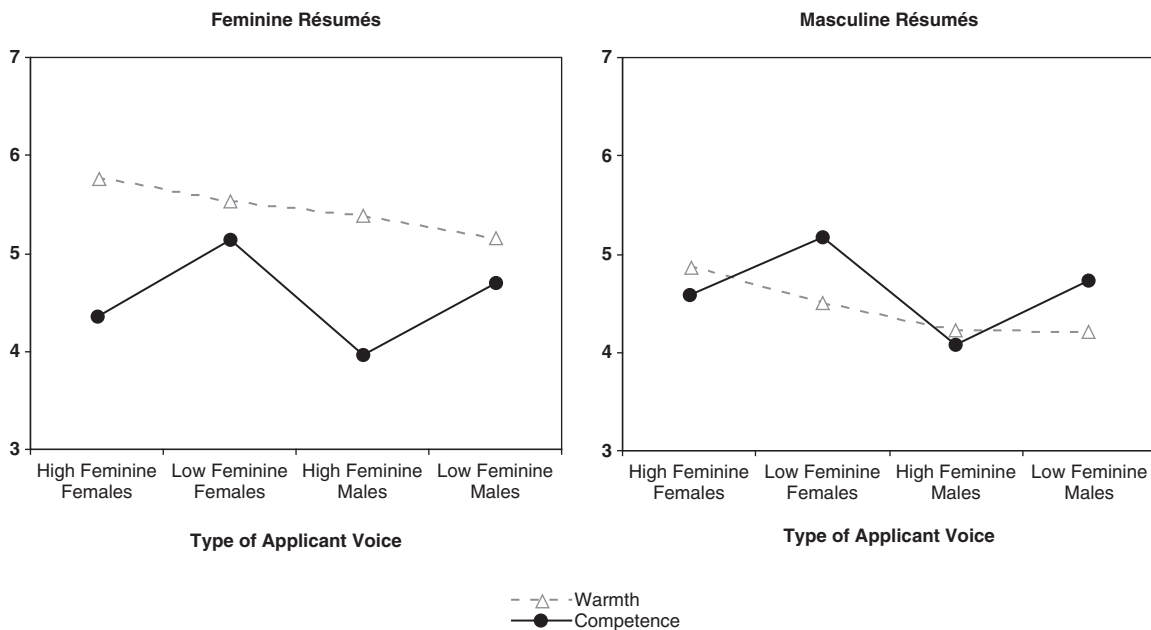
Figure 1 Oral presentation: Composite ratings of warmth and competence on the basis of voice category and vocal femininity for each résumé type.

Importantly, these main effects were moderated by higher order interactions. First, the vocal femininity by judgment dimension interaction was highly significant, $F(1, 61) = 76.85$, $p < .0001$. For judgments of competence, applicants with low feminine voices (within gender) were rated as significantly more competent than those with high feminine voices, $F(1, 61) = 84.50$, $p < .0001$; whereas for judgments of warmth, applicants with high feminine voices were rated as significantly warmer than those with low feminine voices, $F(1, 61) = 8.43$, $p < .01$. Importantly, the magnitude of the vocal femininity effect was much greater on competence than warmth judgments. Second, trait dimension interacted with résumé type, $F(1, 61) = 55.11$, $p < .0001$, such that on the warmth dimension, applicants with feminine résumés were judged as significantly warmer than applicants with masculine résumés, $F(1, 61) = 91.72$, $p < .0001$, whereas these résumé types did not differentially affect judgments on the competence dimension, $F(1, 61) = 1.02$, $p > .30$.

While the vocal femininity by judgment dimension interaction suggests stereotyping on the basis of within-gender vocal femininity, these data, like the text presentation data, offer no evidence of stereotyping based exclusively on gender category, in that the gender category by judgmental dimension interaction was not significant.[6]

Discussion

Our goal in this first experiment was to examine the impact of vocal femininity on stereotypic judgments in the presence of other individuating information. To do that, we created résumés that differed in the degree to which they were sex-typed. Participants then rated each supposed job applicant on the dimensions of warmth and competence. In the text-only condition, the résumé and gender of target were presented. In the audio condition, this information (résumé and gender) was presented orally by target individuals who varied systematically in their

relative vocal femininity. Somewhat surprisingly, across both the text-only and audio conditions, there was a gender main effect such that female targets were rated more highly on both competence and warmth than were male targets. This main effect of gender, across dimension, is inconsistent with traditional gender stereotypes, which portray females as warmer but less competent than males. This may reflect our participants' goals of not derogating women in hiring decisions.

From our perspective, the most important result was found in the audio condition; despite the absence of traditional gender stereotypes, there was a highly significant vocal femininity effect even in the presence of substantial individuating information. In terms of the specific nature of vocal femininity effects, we had presumed it plausible that vocal cues would impact both warmth and competence equally. However, this was not the case. Rather, the effect of vocal femininity was most pronounced on judgments of competence such that across both male and female targets and across résumé type, more feminine sounding targets were judged as less competent. In fact, as reviewed in the introduction, there is literature on voice that might have led us to be more confident in predicting that vocal cues would have a particularly dominant influence on competence. Interestingly, the résumé effects were somewhat inconsistent across the text-only and audio conditions. In the former, targets with more feminine résumés were judged as warmer and less competent. In the audio condition, however, where vocal femininity had a pronounced impact on competence judgments, résumé effects were found only on the warmth judgments. In some sense, then, we might say that vocal femininity came to dominate in the audio condition, wiping out résumé effects on the competence dimension. This difference speaks to the power of these vocal femininity effects; they persist even in the presence of what is otherwise strong individuating information that obviously was relevant to both warmth and competence judgments when voice was not available.

Even though in the introduction we had hypothesized that vocal cues may have a more powerful influence on competence than warmth, this hypothesis was made tentatively due to a lack of any direct evidence that could inform us about this. Hence, the next crucial step was to replicate these results. This was part of our motivation in conducting a second experiment. Additionally, Experiment 2 allowed us to test more stringently the robustness of these novel findings and clarify some of the issues raised in Experiment 1. First, because we only included voices that were rated extremely on vocal femininity for their gender, it is unclear whether gender stereotyping on the basis of vocal femininity would emerge with voices that varied

more continuously in their femininity within gender. Second, it is possible that the absence of gender category stereotyping was due to the particular experimental context (i.e., a job hiring scenario) in which ratings were made. In other words, this context may have been one in which participants were sensitive to issues of gender discrimination. It therefore seemed likely that in other judgment contexts we might find evidence for both gender-based stereotyping and also stereotyping on the basis of vocal femininity. Third, we wanted to extend these results by providing individuating information that was directly relevant to the two judgmental dimensions. This would allow for a more controlled and direct test of the obtained vocal femininity effects in the presence of individuating information. Finally, as reviewed in the introduction, the work on babyfacedness (e.g., Berry & McArthur, 1986) has shown considerable overlap between babyish and feminine/masculine facial features. Based on this, we speculated that to the extent that vocal femininity cues are also correlated with babyish vocal cues, it should explain why vocal femininity is more strongly associated with competence than warmth; competence is the main dimension that distinguishes adults from babies. To test this, we examined whether controlling for vocal babyishness would reduce the effect of vocal femininity on competence.

EXPERIMENT 2: OVERVIEW

For this experiment we modified a version of Locksley et al.'s (1982) paradigm (see also Beckett & Park, 1995) where participants were asked to predict how targets would behave in a particular scenario after learning about targets' behaviors in other diagnostically relevant scenarios (individuating information). All the behaviors were directly related to either the dimension of competence or warmth. Furthermore, the individuating information was presented by target voices that varied both in gender category and (within-gender) vocal femininity.

In the first experiment gender discrimination could have been a salient and sensitive issue because the job hiring context was made salient by the instructions and such contexts have long been associated with such biases. Hence, participants may have been particularly motivated to avoid using stereotypes on the basis of category membership. In contrast, gender discrimination should be a less salient concern in the current experiment because the experimental task is one that does not make gender biases salient.

Another reason for the lack of gender-based stereotyping in Experiment 1 may have been due to "shifting standards" (Biernat, Manis, & Nelson, 1991). According to this theory, subjective rating scales—such as the

Likert-type response scales—are susceptible to biases resulting from raters using different judgment standards to evaluate members of each gender category. For example, even if participants in our experiment gave both a female and a male applicant an identical 6 on assertiveness (higher numbers indicating more assertiveness), the threshold for judging the degree of assertiveness for these two applicants may have been different. Specifically, because assertiveness is stereotypically masculine, female applicants may have needed only to appear/sound somewhat assertive for them to be rated a 6, whereas for male applicants they may have needed to appear/sound extremely assertive for them to be rated a 6. In Experiment 2 we asked participants to predict the likelihood of specific behaviors for each target. Such judgments are likely to be less influenced by "shifting standards."

As a result of these modifications, we expected gender category to impact judgments in Experiment 2 such that on average, males would be rated as more likely to behave competently and less likely to behave warmly than females. However, given past work on feature effects within categories (e.g., Blair et al., 2002; Ko et al., 2006) coupled with our findings from Experiment 1, we predicted that vocal femininity would be a potent enough cue to affect competence judgments over and above any effects of gender category and individuating information. Finally, the effect of vocal femininity on competence might be reduced once we control for vocal babyishness.

EXPERIMENT 2

In this experiment, participants listened to recordings of a series of target persons reporting how they had behaved in four previous scenarios. Specifically, for each target person, participants had to predict how that person would behave in a fifth scenario, based on the information they had heard about the target's behaviors in the previous four scenarios. For half of the participants, the behaviors (and scenarios) were relevant to the competence of the target person, for the other half the behaviors were relevant to the warmth of the target person. Within participants, targets varied in their gender (as indicated by their voice), their vocal femininity (varying within gender), and the degree to which they acted in a competent (or warm) manner in the previous four scenarios.

STIMULUS MATERIALS

Competence and Warmth Behavioral Scenarios

We developed five scenarios in which the behavior of stimulus persons would be indicative of their competence (competence condition) and another five scenarios in which the behavior of stimulus persons would be indicative of their warmth (warmth condition). For example, in one of the warmth scenarios, the target person could either decide to work on a problem collaboratively with another person (warmth behavior) or alone (cold behavior). And in one of the competence scenarios, the target person was either successful at completing a computer puzzle within the allotted time limit (competent behavior) or unsuccessful (incompetent behavior).

Recorded Participants

In all, 71 students (35 female and 36 male) at the University of Groningen were recorded in exchange for monetary compensation. All were native Dutch speakers with an average age of 23 years ($SD = 2.90$).

We recorded the participants in a quiet room using a head-worn microphone. Each participant was asked to read aloud the *Rainbow Passage* and the 16 possible outcome combinations of competent and incompetent responses in the competence condition's first four scenarios and the 16 possible outcome combinations of warmth and cold responses in the warmth condition's first four scenarios.

Voice Judgments, Scaling, and Results

In all, 60 undergraduate students (44 female and 16 male) at the University of Groningen provided vocal femininity and babyishness judgments of the recorded voices, in exchange for partial fulfillment of course credit or monetary compensation. The procedure and methods for the judgments and scaling of these voices were similar to those described in Experiment 1 earlier, with the only difference being that participants listened to the voices two times. The first time, they rated the voices on femininity and the second time on babyishness. The babyishness scale ranged from 1 (*not at all babyish*) to 7 (*very babyish*). For the female voices, the mean femininity and babyishness scores were 4.52 ($SD = 0.89$) and 3.61 ($SD = 1.04$), respectively, and the reliabilities of their means were both .94. For the male voices, the mean femininity and babyishness scores were 3.18 ($SD = 0.93$) and 3.25 ($SD = 0.90$), respectively, and the reliabilities of their means were .93 and .91. Again, the female voices were rated as significantly more feminine than the male voices, $F(1, 69) = 38.63$, $p < .0001$. However, the gender difference in babyishness was not significant, $F(1, 69) = 2.36$, $p > .12$. The correlation between vocal femininity and babyishness was .54 ($p < .001$) for female voices and .89 ($p < .0001$) for male voices. Hence, as expected, babyishness showed considerable overlap with femininity. On the basis of the vocal

femininity scaling, we chose a set of 16 female voices and a set of 16 male voices that maximized the variance on femininity for the main part of Experiment 2. The mean femininity ratings for these voices (and standard deviations) were very similar to those obtained for the full set of voices.

EXPERIMENT 2: BEHAVIORAL JUDGMENTS

Method

Participants. In all, 66 undergraduate students at the University of Groningen participated for partial fulfillment of course credit or monetary compensation. Half of them (25 female and 8 male) were randomly assigned to the competence condition and the other half (22 female and 11 male) were randomly assigned to the warmth condition.

Procedure. On arrival, participants were seated in front of individual computers with headphones. All instructions were presented via the computer. The instructions informed the participants that the experiment was designed to investigate whether knowledge of a person's prior behavior enabled one to accurately predict subsequent behavior. Accordingly, participants were told that they would be listening to 32 speakers who participated in a series of five experiments (i.e., scenarios) intended to measure either, depending on the condition, competent behavior or warmth behavior. Specifically, the speakers would be relating how they acted in the first four scenarios and the participant's task was to predict how the speakers acted in the fifth. For example, a speaker in the competence condition who behaved competently in the first two scenarios and incompetently in the latter two scenarios would have said the Dutch equivalent to "In Experiment 1, I successfully completed the computer puzzle in the allotted time limit. In Experiment 2, I was able see the items' survival values and correctly rank ordered them. In Experiment 3, my observations led me to incorrectly assign the workers to the problems. In Experiment 4, I could not recall the sculpture so was unable to recreate it with the Legos." The participants were told that their predictions would be compared with the speakers' actual behavior in the fifth scenario, thus encouraging accurate predictions. Following these instructions, participants read the description of each of the five scenarios relevant to their condition and for each were concretely told the behavior that was considered competent (warm) and the behavior that was considered incompetent (cold). Participants then listened to all 32 male and female speakers and for each speaker they had to estimate the probability (0% to 100%) that the

speaker behaved competently (warmly) in the fifth scenario. In each condition, the four scenarios in which prior behaviors were provided yielded 16 possible outcome combinations (e.g., competent in all four; competent in 1, 2, and 3, but not 4; competent in 1, 2, and 4 but not 3; etc.) and prior to the experiment, per condition, the 16 stimulus speakers in each gender category were randomly paired with the 16 outcome combinations, thus unconfounding gender of speakers and vocal femininity. The order of the speakers was randomized for each participant. After completing the ratings, participants were debriefed and dismissed.

Analysis and Results

To analyze the resulting data, we adopted a multi-level modeling approach, estimating the effects of prior behavior, gender category, and vocal femininity on predictions of behavior in the fifth scenario, within each participant. Then, across participants, we examined the impact of each one of these factors (prior behaviors, gender category, and vocal femininity) on behavioral predictions separately for the competence and warmth (between participants) conditions (for a similar analytic approach, see Blair, Chapleau, & Judd, 2005).

We first estimated the effect of individuating information alone on the competence and warmth judgments. For this purpose, we estimated a regression model for each participant where the likelihood of competent (warm) behavior in the fifth scenario for the 32 speakers was regressed onto the proportion of competent (warm) behavior(s) in the first four scenarios. Next, in order to examine the critical question of whether gender category of the speaker and vocal femininity affected competence and warmth judgments, over and above the individuating information given about each target person in the four scenarios, we estimated a second model for each participant, where the likelihood of competent (warm) behavior in the fifth scenario was predicted from gender category (contrast coded as 1 if female and –1 if male), vocal femininity, and the proportion of competent (warm) behaviors in the first four scenarios. These analyses gave us four slopes for each participant, one from the first model and three from the second. Across participants, we then tested whether mean slopes differed from zero and also whether the mean slopes differed between the competence and warmth conditions.

Presented in the top half of Table 2 are the mean within-participant slopes from the first model, where individuating information was the sole predictor of the likelihood ratings. The slopes were significantly different from zero in the expected direction. That is, for the competence ratings, the more competent behaviors the target persons had manifested in the previous four

TABLE 2: Experiment 2: Mean Slopes Predicting Likelihood of Behavior in the Fifth Experiment From Gender Category, Vocal Femininity, and Behavioral Proportions in the First Four Experiments (Individuating Information), Broken Down by Condition

	Condition	
	Competence	Warmth
Model 1: Individuating information as sole predictor		
Individuating information slopes	0.62** (.70)	.69** (.78)
	Condition	
	Competence	Warmth
Model 2: Category, vocal femininity, and and individuating information as predictors		
Partial slopes for		
Vocal femininity	−1.46** (−.08)	−.23 (−.01)
Category	−1.81* (−.08)	.58 (.03)
Individuating information	0.64** (.73)	.68** (.77)

NOTE: For category slopes, negative numbers indicate male voices judged as more likely than female voices.
*Slope significantly different from zero, $p < .01$. **Slope significantly different from zero, $p < .001$.

scenarios, the more likely they were predicted to behave in a competent manner in the fifth (mean $R^2 = .51$). Likewise for the warmth ratings, the more warm behaviors the target persons had manifested in the first four scenarios, the more likely they were predicted to behave warmly in the fifth (mean $R^2 = .60$). Furthermore, the competence slopes did not differ from the warmth slopes, $F(1, 64) = 1.74$, $p < .20$, showing that the strong influence of individuating information did not differ by condition.

Presented in bottom half of Table 2 are the means of the partial slopes from the second model, broken down by condition. All partial slopes in the competence condition were significantly different from zero and in the expected direction. That is, the more competent behaviors that targets manifested in the previous four scenarios, the more likely they were predicted to behave competently in the fifth; male targets were predicted to behave more competently, controlling for their individuating information (i.e., behavior in the previous four scenarios); and the less feminine the voice, the more competent the prediction, controlling both for individuating information and gender category. On the other hand, in the warmth condition, only the partial slopes for individuating information were significantly different from zero.

Testing the between-condition differences in these slopes revealed that the effect of gender category on judgments in the fifth scenario differed significantly by condition, $F(1, 64) = 7.27$, $p < .01$, such that in the competence condition, male speakers were rated as more likely to behave competently than female speakers in the fifth scenario, controlling for individuating information and vocal femininity, whereas gender category had no

effect on the likelihood ratings in the warmth judgments. Additionally and importantly, the partial slopes for vocal femininity differed significantly between the two conditions, $F(1, 64) = 4.70$, $p < .05$, such that in the competence condition, less vocal femininity was associated with higher competence predictions, controlling for gender category and individuating information, whereas vocal femininity had no effect on warmth behavioral predictions.

Having replicated the vocal femininity effect on competence that we found in Experiment 1, our final analysis examined whether this vocal femininity effect would be reduced once we control for vocal babyishness. Accordingly, we regressed the competence behavioral predictions on vocal femininity, vocal babyishness, contrast-coded gender category, and proportion of competent behaviors. As can be seen in Table 3, adding babyishness as a predictor to the model did not reduce the significant impact of category or individuating information. Importantly however, our predictions were supported in that including babyishness did reduce the influence of vocal femininity such that vocal femininity was no longer a significant predictor of competence. This result is consistent with the suggestion that vocal femininity effects on competence are largely mediated by babyishness.

Discussion

Experiment 2 allowed a more rigorous and controlled test of the key findings and predictions from Experiment 1. Specifically, we used a well-established experimental paradigm first developed by Locksley et al.

TABLE 3: Mean Slopes Predicting Likelihood of Competence Behavior in the Fifth Experiment From Vocal Femininity, Vocal Babyishness, Category, and Competence Behavioral Proportions

Partial slopes for	Competence
Vocal femininity	0.29 (.01)
Vocal babyishness	–2.20* (–.09)
Category	–2.56** (.12)
Individuating information	0.62** (.70)

NOTE: For category slopes, negative numbers indicate male voices judged as more likely than female voices.
*Slope significantly different from zero, $p < .01$. **Slope significantly different from zero, $p < .001$.

(1982, Experiment 2), where diagnostic individuating information has consistently been shown to exert a strong influence on judgments. Over and above the strength of the individuating information, this experiment provides an even stronger test of the vocal femininity effects because we used a more objective dependent measure (i.e., behavioral predictions). For the voice stimuli, we selected voices that varied continuously on vocal femininity (within gender), hence, countering the possibility that the vocal femininity effects in the first experiment resulted from the use of voices that were relatively extreme in femininity (again, within gender).

Even under these more stringent conditions, the vocal femininity results from Experiment 1 were replicated. That is, even with the substantial influence of individuating information, both gender category and within-category vocal femininity of the speakers affected behavioral predictions in the competence condition, whereas in the warmth condition, no such effects were observed. Specifically, speakers on the basis of their voice category differed in how competently they behaved in the fifth scenario. Additionally and most important, vocal femininity affected competence judgments, over and above gender category. Furthermore, as hypothesized, the facet of femininity/masculinity captured by vocal femininity did have considerable overlap with vocal cues that signal babyishness. Consequently, just as adults and babies are (fundamentally) distinct on competence, but not necessarily on warmth, feminine/masculine voices likewise triggered strong associations with competence but not warmth judgments.

In the first experiment we found that the individuating information contained in the résumés had a greater impact on warmth judgments than on competence judgments. That effect was not replicated here, since the individuating information conveyed by behaviors in the prior four scenarios had a large impact on both competence and warmth predictions in the fifth. This is not surprising since in this study the individuating information was directly diagnostic of each dimension. In other words, it is as expected that competence-relevant individuating information influenced competence judgments just as strongly as warmth-relevant individuating information influenced warmth judgment. Another difference between the two experiments was the gender category effect that was absent in the first. As discussed earlier, there are two factors that could have contributed to this. First, the job hiring context of Experiment 1 may have made participants wary of appearing biased, and as a result, they may have curbed their use of category-based judgments. The task in Experiment 2, on the other hand, should not have made such concerns salient. Hence, the participants in the second experiment may not have been so wary of using category-based judgments. Second, the dependent measure in Experiment 1 may have been influenced by shifting standards of judgments, whereas the behavioral predictions in Experiment 2 should have been more resistant to such judgmental adjustments as a function of target gender category.

It is possible that the effects of individuating information on the two dimensions are context dependent. That is, when participants think they are in job hiring contexts (as in the first experiment), they may weigh competence more heavily than warmth to assess applicant suitability (e.g., Bertrand & Hallock, 2001; Blau & Kahn, 2000; Eagly & Karau, 2002; Nesbitt & Penn, 2000; Powell, Butterfield, & Parent, 2002; Schein, 1975, 1978). Given that recruiters seem to place a great deal of emphasis on interviews in order to make their final hiring decisions, it is plausible that the physical cues provided by such contexts—which are absent in written résumés—are perceived as particularly diagnostic of job success. Consequently, participants may rely heavily on applicants' physical cues to make competence judgments, especially as these cues seem to be strongly associated with maturity. In the current work, we found vocal cues to be the physical cue of importance in judging applicants' competence. This has important implications for job interviews, especially as telephone interviews may be the most common way that recruiters screen applicants. Obviously we do not know if vocal cues would continue to have such an impact on competence judgments in situations where applicants' other physical cues are concurrently available, hence future research should investigate this issue. However, their potent effects suggest that inferences from vocal cues are not trivial, even in the presence of individuating information and other perceptual cues.

Conclusion

Taken together, these results provide an exciting demonstration that something as subtle and yet as prevalent

as vocal cues continue to impact judgments in the face of other competing information. Furthermore, our research in auditory stereotyping provides a richer perspective of the stereotyping literature by shifting what has commonly been its singular focus on visual stereotyping to include examination of auditory-based judgments. This shift allows for investigations of stereotyping from multiple angles. A significant benefit to this is that as we acquire more insight into auditory stereotyping, we may find that there are not only commonalities with the more established visual stereotyping processes, but also intriguing differences showing how vocal cues may have unique effects on certain stereotypic judgments. In the current research, there was certainly a hint of this in the way that vocal cues affected the judgmental dimension of competence while having little influence on the dimension of warmth. In sum, research in auditory stereotyping may yield findings that can be used to generate new theories on auditory-based perception and add to the existing theories on stereotyping.

In the larger scheme of things, there is vast potential for application of our research. Due to the incredible malleability and richness of the voice, learning to modify it, as the situation requires, would be a relatively effortless undertaking that returns rich benefits. Indeed, a recent article in *The New York Times* (Jaret, 2005) detailed a major increase in the number of people who seek so-called voice makeovers to change the impressions that they are making. These people are no longer just actors, singers, and famous people, as was the case in the past: Now they range from marketing consultants from San Jose to second-grade teachers in Manhattan. Clearly, their everyday experiences have taught them just how important the voice may be in the impressions that they are making. In the end it may not only be who you are, what you look like, and what you have accomplished, but also how you sound that affects the impressions formed of you.

NOTES

1. An English version of all experimental materials can be obtained by contacting the first author.

2. The participants also rated the résumés on valence using a 1 (*very negative*) to 7 (*very positive*) scale. The valence mean ratings for the chosen résumés were as follows: feminine résumés were 5.18; masculine résumés were 4.74. Pairwise comparisons yielded differences between the feminine and masculine résumés, $t(1, 15) = 2.74$, $p < .05$.

3. In our previous work (Ko, Judd, & Blair, 2006) we had had separate participants scale voices on masculinity. The extremely high negative correlation of the resulting masculinity ratings with the femininity ratings ($r = -.92$) led us to conclude that vocal femininity and masculinity were opposite ends of the same dimension. Accordingly, voices low on femininity are voices high on masculinity, and vice versa, voices high on femininity are voices low on masculinity.

4. The reliabilities of these composites could be estimated either across targets for each participant or across participants for each target.

We did the latter using the text presentation data. The average reliability for competence items was .69 and for warmth items was .56.

5. We found no participant gender effects across the two experiments, and hence, this will not be discussed further.

6. The intercorrelations between competence and warmth ranged from $r = -.23$, $p = .07$ to $r = .47$, $p < .05$. We did not observe any systematic pattern of difference in their magnitude as a function of résumé sex-type, category, vocal femininity, or résumé presentation.

REFERENCES

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

Bakan, D. (1966). *The duality of human existence*. Chicago: Rand McNally.

Beckett, N. E., & Park, B. (1995). Use of category versus individuating information: Making base-rates salient. *Personality and Social Psychology Bulletin*, 21, 21-31.

Berry, D. S., & McArthur, L. Z. (1986). Perceiving character in faces—The impact of age-related craniofacial changes on social-perception. *Psychological Bulletin*, 100, 3-18.

Bertrand, M., & Hallock, K. F. (2001). The gender gap in top corporate jobs. *Industrial and Labor Relations Review*, 55, 3-21.

Biernat, M., & Manis, M. (1994). Shifting standards and stereotype-based judgments. *Journal of Personality and Social Psychology*, 66, 5-20.

Biernat, M, Manis, M., & Nelson, T. E. (1991). Stereotypes and standards of judgment. *Journal of Personality and Social Psychology, 60*, 485-499.

Blair, I. V., Judd, C. M., & Chapleau, K. M. (2004). The influence of Afrocentric facial features in criminal sentencing. *Psychological Science*, 15, 674-679.

Blair, I. V., Judd, C. M., & Fallman, J. L. (2004). The automaticity of race and Afrocentric facial features in social judgments. *Journal of Personality and Social Psychology*, 87, 763-778.

Blair, I. V., Judd, C. M., Sadler, M. S., & Jenkins, C. (2002). The role of Afrocentric features in person perception: Judging by features and categories. *Journal of Personality and Social Psychology*, 83, 5-25.

Blau, F. D., & Kahn, L. M. (2000). Gender differences in pay. *Journal of Economic Perspectives*, 14, 75-99.

Brown, B. L., Strong, W. J., & Rencher, A. C. (1973). Perceptions of personality from speech: Effects of manipulations of speech from acoustical parameters. *Journal of Acoustical Society*, 54, 29-35.

Cejka, M. A., & Eagly, A. H. (1999). Gender-stereotypic images of occupations correspond to the sex segregation of employment. *Personality and Social Psychology Bulletin*, 25, 413-423.

Davison, H. K., & Burke, M. J. (2000). Sex discrimination in simulated employment contexts: A meta-analytic investigation. *Journal of Vocational Behavior*, 56, 225-248.

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

DeGroot, T., & Motowidlo, S. J. (1999). Why visual and vocal interview cues can affect interviewers' judgments and predict job performance. *Journal of Applied Psychology*, 84, 986-993.

Eagly, A. H. (1987). *Sex differences in social behavior: A social-role interpretation*. Hillsdale, NJ: Lawrence Erlbaum.

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, 109, 573-598.

Eagly, A. H., & Mladinic, A. (1989). Gender stereotypes and attitudes toward women and men. *Personality and Social Psychology Bulletin*, 15, 543-558.

Eagly, A. H., & Steffen, V. J. (1984). Gender stereotypes stem from the distribution of women and men into social roles. *Journal of Personality and Social Psychology*, 46, 735-754.

Eberhardt, J. L., Davies, P. G., Purdie-Vaughns, V. J., & Johnson, S. L. (2006). Looking deathworthy: Perceived stereotypicality of Black defendants predicts capital-sentencing outcomes. *Psychological Science*, 17, 383-386.

Fairbanks, G. (1940). *Voice and articulation drillbook*. New York: Harper & Brothers.

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*, 878-902.

Fiske, S. T., Xu, J., Cuddy, A. J. C., & Glick, P. (1999). (Dis)respecting versus (dis)liking: Status and interdependence predict ambivalent stereotypes of competence and warmth. *Journal of Social Issues*, *55*, 473-491.

Friedman, H., & Zebrowitz, L. A. (1992). The contribution of typical sex-differences in facial maturity to sex-role stereotypes. *Personality and Social Psychology Bulletin*, *18*, 430-438.

Glick, P., Wilk, K., & Perreault, M. (1995). Images of occupations—Components of gender and status in occupational stereotypes. *Sex Roles*, *32*, 565-582.

Glick, P., Zion, C., & Nelson, C. (1988). What mediates sex discrimination in hiring decisions? *Journal of Personality and Social Psychology*, *55*, 178-186.

Heilman, M. E. (2001). Description and prescription: How gender stereotypes prevent women's ascent up the organizational ladder. *Journal of Social Issues*, *57*, 657-674.

Jaret, P. (2005, July 21). My voice has got to go. *The New York Times*. Retrieved April 29, 2006, from http://www.nytimes.com/2005/07/21/fashion/thursdaystyles/21voice.html?ei=5090&en=8db17c665c6700a8&ex=1279598400&partner=rssuserland&emc=rss&pagewanted=all

Judd, C. M., James-Hawkins, L., Yzerbyt, V., & Kashima, Y. (2005). Fundamental dimensions of social judgment: Understanding the relations between judgments of competence and warmth. *Journal of Personality and Social Psychology*, *89*, 899-913.

Judd, C. M., & McClelland, G. H. (1998). Measurement. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., pp. 180-232). New York: McGraw-Hill.

Ko, S. J., Judd, C. M., & Blair, I. V. (2006). What the voice reveals: Within- and between-category stereotyping on the basis of voice. *Personality and Social Psychology Bulletin*, *32*, 806-819.

Ko, S. J., Muller, D., Judd, C. M., & Stapel, D. A. (2008). Sneaking in through the back door: How category-based stereotype suppression leads to rebound in feature-based effects. *Journal of Experimental Social Psychology*, *44*, 833-839.

Livingston, R. W., & Brewer, M. B. (2002). What are we really priming? Cue-based versus category-based processing of facial stimuli. *Journal of Personality and Social Psychology*, *82*, 5-18.

Locksley, A., Borgida, E., Brekke, N., & Hepburn, C. (1980). Sex stereotypes and social judgment. *Journal of Personality and Social Psychology*, *39*, 821-831.

Locksley, A., Hepburn, C., & Ortiz, V. (1982). Social stereotypes and judgments of individuals: An instance of the base-rate fallacy. *Journal of Experimental Social Psychology*, *18*, 23-42.

Macrae, C. N., Bodenhausen, G. V., Milne, A. B., & Jetten, J. (1994). Out of mind but back in sight—Stereotypes on the rebound. *Journal of Personality and Social Psychology*, *67*, 808-817.

Maddox, K. B., & Gray, S. A. (2002). Cognitive representations of Black Americans: Reexploring the role of skin tone. *Personality and Social Psychology Bulletin*, *28*, 250-259.

Miller, N., Maruyama, G., Beaber, R. J., & Valone, K. (1976). Speed of speech and persuasion. *Journal of Personality and Social Psychology*, *34*, 615-624.

Montepare, J. M., & Zebrowitz, L. (1998). Person perception comes of age: The salience and significance of age in social judgments. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 30, pp. 93-161). San Diego, CA: Academic Press.

Nesbitt, M. N., & Penn, N. E. (2000). Gender stereotyping after thirty years: A replication of Rozenkrantz, et al. (1968). *Psychological Reports*, *87*, 493-511.

Powell, G. N., Butterfield, D. A., & Parent, J. D. (2002). Gender and managerial stereotypes: Have times changed? *Journal of Management*, *28*, 177-193.

Rosenberg, S., Nelson, C., & Vivekananthan, P. S. (1968). A multidimensional approach to the structure of personality impressions. *Journal of Personality and Social Psychology*, *9*, 283-294.

Schein, V. E. (1975). Relations between sex role stereotypes and requisite management characteristics among female mangers. *Journal of Applied Psychology*, *60*, 340-344.

Schein, V. E. (1978). Sex role stereotyping, ability and performance: Prior research and new directions. *Personnel Psychology*, *31*, 259-268.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*, 420-428.

Smith, B. L., Brown, B. L., Strong, W. J., & Rencher, A. C. (1975). Effects of speech rate on personality. *Language and Speech*, *18*, 145-152.

Spence, J. T. (1993). Gender-related traits and gender ideology—Evidence for a multifactorial theory. *Journal of Personality and Social Psychology*, *64*, 624-635.

Tusing, K. J., & Dillard, J. P. (2000). The sound of dominance: Vocal precursors of perceived dominance during interpersonal influence. *Human Communication Research*, *26*, 148-171.

Zebrowitz, L. A., & Montepare, J. M. (2005). Appearance DOES matter. *Science*, *308*, 1565-1566.