A Reader in Sociophonetics

edited by
Dennis R. Preston
Nancy Niedzielski

(2010)

De Gruyter Mouton

pp 327-356

Chapter 14

Sound Judgments: Perception of Indexical Features in Children's Speech

Paul Foulkes, University of York; Gerard Docherty and Ghada Khattab, Newcastle University; Malcah Yaeger-Dror, University of Arizona

1. Introduction

One of the defining features of human language is that it displays systematic variation at all levels of structure, from syntax to fine-grained features of pronunciation. Certain aspects of this variation result from biological differences across speakers. One example of biologically-constrained variation is the markedly different levels of fundamental frequency typical of men, women, and children, which result to a large extent from gross differences in laryngeal anatomy and physiology. Variation may also derive from learned patterns of behavior, acquired as a consequence of a speaker's regional, social, linguistic, and cultural background. Regional accent is a clear case in point.

When we speak we therefore offer a wealth of information about ourselves through the linguistic and phonetic alternatives we use. Listeners can and do take notice of such alternatives. Several studies have investigated which variable cues listeners can identify, how they are identified, and what interpretations listeners make of them. It has been shown, for example, that cues related to speaker gender, or to individual talkers, can affect linguistic processing such as lexical identification or phoneme categorization (e.g., Johnson 1997, Strand 1999, Hawkins and Smith 2001). Sociolinguistic studies, meanwhile, have established that listeners evaluate variants (positively or negatively), linking them to aspects of personality such as intelligence and friendliness (e.g., Giles and Powesland 1975). Furthermore, listeners can identify, with varying degrees of accuracy, aspects of a talker's social, ethnic and regional background (e.g., van Bezooijen and Gooskens 1999; Purnell, Idsardi, and Baugh 1999; see further Thomas 2002 for an excellent summary of speech perception work which bears on sociolinguistic issues).

In this chapter we describe an experiment in which listeners were asked to identify the sex of children from their speech. The study has two broad aims. First, it seeks to enhance our understanding of the range of cues used by listeners in performing the task of gender identification. Although a number of studies have addressed this issue previously (reviewed subsequently), results have been somewhat inconsistent and at times vague. Secondly, our specific interest is in the role played by fine-grained phonetic variants in listeners' responses. It is well known from sociolinguistic studies that segmental features may vary quantitatively within a community. A particular form may therefore be indexical of a social group. For instance, men might use statistically more of a particular variant than women do in the same type of interactional speech style. However, little work has been carried out to assess whether listeners show any awareness of such statistical associations between phonetic forms and social categories.

We begin with a brief review of previous work on identifying speaker sex, and of sex-correlated variation in speech. In Section 3 we then outline the dialect and sociolinguistic variables of interest for our study. Section 4 explains the experimental method we adopted, and the results are discussed in Section 5. The final section summarizes the findings and identifies opportunities for further work.

2. Listener identification of speaker sex—Previous studies and possible cues

We might expect that judging the sex of a speaker is relatively straightforward, at least for adult talkers engaged in everyday interaction. This perhaps explains why few studies have tested listeners' ability to distinguish the sex of adult male and female talkers. One of the few studies to include a formal identification test of speaker sex (as part of a larger project) reports a 100% success rate (Krauss, Freyberg, and Morsella 2002: 621).

For adults, as noted earlier, a particularly robust cue to speaker sex is provided by the fundamental frequency (f0) of the speaker's voice. Based on analysis largely of western European languages, the average f0 for male speakers is around 120 Hz while that for females is around 220 Hz (e.g., Fant 1956). Klatt and Klatt (1990) estimate that female f0 averages around 1.7 times that of males.

However, f0 is neither an infallible nor an invariant cue to speaker sex. First of all, there is considerable overlap in the f0 ranges used by adult males and females, such that a high pitched male voice may be mistaken for a low pitched female voice, or vice versa. For f0 in normal conversation Fant (1956) identifies the maximal male range as 50–250Hz, while that for women overlaps the male range at 120–480 Hz. Künzel (1987) presents data from 100 Germans which indicate that around 35% of males have an average f0 higher than the female baseline of 120 Hz. Moreover, f0 is subject to variation in response to many factors. Analysis of speech in contexts other than regular conversation shows, for example, that male f0 can be much higher under stress (Boss 1996), when a speaker attempts to counter ambient noise (e.g., Lane and Tranel 1971), and in telephone speech or when reading aloud (Hirson, French, and Howard 1995). It also appears that average male f0 may deviate markedly from the often-cited mean of 120 Hz when we consider different languages and non-standard dialects. French and Harrison (2005), for instance, report a mean of 105 Hz for 22 Caribbean males in Birmingham, UK, while the average for speakers of Urdu has been reported to be as high as 186 Hz (Peter French, personal communication).

Although f0 may be an obvious cue in many instances, it does not always provide unambiguous information about a speaker's sex. What is more, in the case of prepubescents, f0 may not be a helpful cue at all. The gross differences we can observe in f0 across males and females result from the physiological changes which occur in male voices at the onset of puberty (the "breaking" of the voice; Mackenzie Beck 1997). Males have a lower f0 because their vocal folds are larger and more massive, and thus they vibrate more slowly than those of females. With children there are no such major physiological differences (although there is some evidence that small physical differences in vocal tract anatomy emerge well before puberty, e.g., King 1952; Crelin 1973).

Nevertheless, there is abundant evidence that children's speech manifests phonetic differences which listeners can access to identify the sex of the talker. Several studies report response rates well above chance on sex identification tests (reviewed by Perry, Ohde, and Ashmead 2001). However, it is less clear which cue(s) are the most useful for listeners.

Given that f0 is such an important cue for adult talkers, it is understandable that several studies have assessed whether f0 differs for boys and girls, and also whether f0 is used by listeners to judge the sex of child talkers. For example, Günzburger et al. (1987) asked listeners to judge the sex of 17 children, and then recruited a group of blind listeners who rated the three best identified boys and three best identified girls on a number of perceptual scales. The clearest result from the latter part of the study was that the best identified girls were consistently rated as high pitched, while the boys were just as consistently given low pitch ratings. However, the consensus is that average f0 plays at best only a secondary role in listeners' perceptions of sex for child talkers (Weinberg and Bennett 1971; Bennett and Weinberg 1979b; Perry et al. 2001). As expected, f0 has generally been found not to differ systematically or consistently between boys and girls in the way that it does for adults. Presumably it therefore cannot function as a robust cue for speaker sex. Weinberg and Bennett (1971), for instance, report no statistical differences between boys and girls aged 5 and 6, while Lee, Hewlett and Nairn (1995) found significant differences to emerge only from age 12. Perry et al. (2001) identified differences for 16-year-olds (and no differences for 4-, 8- or 12-year-olds), but their focus was on f0 of single vowels extracted from carrier phrases. Some studies in fact report girls to have lower f0 than boys matched for age (e.g., Sachs et al. 1973 for children aged 4 to 14, Günzburger et al. 1987 for Dutch-speaking children aged 7 and 8). By contrast, those studies that have found boys to have lower f0 than girls have usually been based on analysis of short and non-spontaneous materials such as sustained isolated vowels (e.g., Hasek, Singh, and Murry 1980 for ages 7 to 10), or small speaker samples (e.g., Sorenson 1989, reporting significant differences at ages 6, 8, 9 and 10, with three children of each sex per age category). Lieberman (1967) also describes findings from a small study, but his data are nonetheless noteworthy. A ten month old baby boy was recorded with a mean f0 of 390Hz when playing with his mother, but 340 Hz in a similar 20 minute play session with his father. A similar effect was found with a 13-month-old girl (average 390 Hz with the mother and 290 Hz with the father). The implication of these finding is that the children were adjusting their overall f0 level in relation to that of their interlocutor. However, it is clear that f0 does not play the same role in cueing the sex of child talkers as it does for adults.

In dismissing average f0 as an important cue, several studies have suggested that listeners can gain more reliable information from relative vowel formant frequencies and spacing. Perry et al. (2001) found systematic differences in formant values for four-year-olds, with boys giving lower first and second formant values than girls. Children therefore display the same patterns found to a more marked degree for adults. It remains unclear whether these effects are the product of emerging differences invocal tract dimensions, whether boys and girls are imitating the differences that can be observed between adults, or both. Listeners do, however, seem inclined to attribute low F1 and F2 values to boys (Bennett and Weinberg 1979b, Perry et al. 2001).

Few other possible contributors to the identification of talker sex have been identified, and fewer still tested. Sachs et al. (1973: 81) suggest intonation and voice quality as possible cues, commenting also that boys in their study had a "more forceful, definite rhythm," although these parameters are not tested in their study. Günzburger, Bresser, ter Keurs (1987), however, offer some support that intonation patterns might affect listener response in sex identification tasks. Among the perceptual scales used by their blind listener group was "monotonous—melodious." The boys who were most successfully identified were given high "monotonous" ratings, while girls were given high "melodious" ratings. This suggests that a wide intonation range might be taken as an indicator of female speech. Production studies of adult talkers, however, fail to reveal a consistent pattern which would explain why listeners might make this inference. Syrdal's (1996) analysis of the 160 speaker Switchboard corpus, for instance, did find that women had a much wider f0 range than men, whereas Henton's (1989) review of earlier research suggested the opposite general trend. Henton's own study, based on a sample of ten Americans, yielded no significant differences for speaker sex. In comparing these studies it should also be borne in mind that Syrdal's data were collected from telephone calls. Telephone transmission introduces various acoustic and phonetic effects into the speech signal, both via the technical effects of passing the signal through a handset and telephone line, and also because speakers may behave differently when speaking on a telephone (e.g., Moye 1979; Summers et al. 1988; Byrne and Foulkes 2004).

Voice quality covers a wide array of phonetic cues (Laver 1980), only a small number of which have been addressed in production studies where speaker sex has been at issue. Phonatory differences have received the most attention. Breathy phonation has been identified regularly as a characteristic of female speech (e.g., Thorne et al. 1983; Henton and Bladon 1985; Klatt and Klatt 1990; Hillenbrand, Cleveland, and Erickson 1994), although there is often considerable variation within the male and female speaker groups tested. Creak has been attributed both to males (Henton and Bladon 1988 for British English speakers) and females (Syrdal 1996 for American English speakers), which might indicate regionally- or sociolinguistically-governed patterning. Stuart-Smith's (1999) detailed study of Glasgow speech identified both creaky phonation and nasalization as consistent features of male speech, while females were found to use phonation that is more whispery. In spite of such observations on voice quality, however, these do not appear to have been tested in either production or perceptual studies of children's speech.

The Günzburger et al. (1987) study suggested that overall amplitude or intensity of speech may be relevant in understanding listeners' responses to child talkers, with the blind listener group rating boys as "loud" and girls as "soft." Again, however, this seems not to have been the subject of formal testing in perceptual studies. Production studies do offer some support, though.

For example, Markel, Prebor, and Brandt (1972) found that adult males spoke, on average, with a greater intensity than females in interview tasks (the overall male average was 76.1 dB compared with that for females of 69.5 dB).

A final parameter for consideration is articulation rate. Bennett and Weinberg (1979a) considered rate as a possible factor in listeners' judgments, but found no differences in rate when they compared the speech of boys and girls and therefore concluded it would not be likely to affect listeners. A number of studies of adult speech production have also reported no differences in rate (Ryalls et al. 1994, Syrdal 1996, Robb, Maclagan, and Chen 2004). There is, however, some evidence for rate differences in other studies. Byrd (1994) analysed rate for 630 talkers in the TIMIT corpus, using two spoken sentences per subject. She found that men spoke on average 6.2% faster than women (the male mean was 4.69 syllables per second, compared with the female mean of 4.42). Yuan, Liberman, and Cieri (2006) also claim a small but significant effect, again with males speaking faster than females. Their findings are derived from analysis of several large corpora of telephone speech from English and Chinese speakers.

From this brief review of previous studies we can conclude that, while male and female talkers differ on a number of phonetic dimensions, the consistency and systematicity of these features is variable, and the evidence for their value in judgments of speaker sex is sketchy. Moreover, although a number of phonetic features have been addressed in perceptual studies, researchers have not previously assessed the role played by gender-correlated sociolinguistic variables. It is well known that gender patterning in sociolinguistic variable studies is extremely widespread (see the review by Chambers 2003). For example, males (at least in western societies) typically use higher proportions of non-standard variants than females of the same age, social background and community. What remains unclear is whether listeners can recognize the statistical associations between sociolinguistic variants and speaker sex in the way that has been shown for associations with ethnicity, social class, and region.

We move on now to describe the experiment we carried out to test listeners' perceptions of gender-correlated variables. This experiment was originally designed as a pilot study to probe listeners' ability to identify speaker sex from a range of phonetic cues: sociolinguistic variables were our main concern, but we also sought to test the effects of cues such as voice quality and rate, following the predictions made by other researchers which were reviewed previously. We chose to exploit the fact that we had access to a large number of recordings of speech from pre-school children, for whom no gross f0 differences would be expected. With respect to f0, then, we are assuming

that children's voices are in principle androgynous, enabling us to test the contribution of other phonetic factors in listeners' perception of sex.

3. Dialect focus: Sex-correlated patterns in Tyneside voiceless stops

Tyneside is a large conurbation in the north-east of England, with the city of Newcastle upon Tyne as its hub. The dialectology and sociolinguistics of the region have been studied as extensively as any in the British Isles (e.g., Heslop 1892, Pellowe et al. 1972, Pellowe and Jones 1978, Local 1982, Jones-Sargent 1983, Jones 1985). The wide interest in Tyneside is undoubtedly a reflection both of the strong and distinctive cultural identity of the region, and also the singular character of its dialect. The Tyneside dialect, commonly referred to as Geordie, is in many respects very different from other non-standard varieties of British English in lexis, syntax, and both segmental and suprasegmental phonology and phonetics (see e.g., Watt and Milroy 1999, Local, Kelly, and Wells 1986, Beal 1993, 2004, and for example sound files www.ncl.ac.uk/necte/ and www.phon.ox.ac.uk/IViE/). The dialect is one that lay listeners find relatively easy to recognize, and its saliency is also testified by copious dialect literature (e.g., Dobson's popular lexicon and grammar *Larn Yersel' Geordie*, and several characters in the Newcastle-based adult comic *Viz*).

Two large empirical projects were executed in Tyneside in the 1990s, the first with adult subjects (Milroy, Milroy, and Docherty 1997) and the second with children and their mothers (Docherty et al. 2002). We draw on the findings and materials of both of these studies here. The adult study is henceforth referred to as the PVC project (an abbreviation of its full title, *Phonological Variation and Change in Contemporary British English*). The child study is abbreviated to ESV (*The Emergence of Structured Variation in the Speech of Tyneside Infants*). The fieldwork design and methods of data collection and analysis have been reported in detail elsewhere (for PVC see Docherty et al. 1997 and for ESV Foulkes et al. 2005), but Table 14.1 provides a summary of the main features of each project.

The work of the first two authors has focused on consonantal variation, with a particular interest in voiceless stops. Auditory and acoustic analysis of voiceless stops reveals rich, complex and often very subtle patterns of variation (see in particular Docherty et al. 1997, Docherty and Foulkes 1999, 2005, Foulkes et al. 2005, Foulkes and Docherty 2006). The plain oral stops [p t k], characteristic of standard English, are also found in Tyneside. However, /p t k/ in Tyneside English may be spirantized, pre-aspirated and/or voiced, while /t/

Table 14.1 Fieldwork Summary for Tyneside Projects

	PVC	ESV
speakers	32 adults	53 children + mothers
age range	15–27, 45–67	1;11—4;1
social class	MC and WC	WC
sex	males and females	males and females
spoken materials	45-minute free conversations in pairs + wordlists	free play sessions including toy-bag and picturebook tasks
location	subjects' homes	subjects' homes
recording media	Sony TCD D-10 Pro II DAT Sennheiser microphone	Sony TCD-D10 Pro II DAT Trantec lapel microphone

can be also realized as a tap or [ɹ]. Most distinctive of all is the range of highly localized variants involving laryngeal constriction. Intersonorant /p t k/ (in *water, winter*) are generally realized with full or partial voicing and a period of creaky phonation either before or after the oral constriction. Wells (1982) transcribes these variants as [p͡ʔ t͡ʔ k͡ʔ] although we prefer [b̰ d̰ g̰] to reflect the fact that such tokens tend to be voiced. These local variants are very salient to the ears of outsiders, and are undoubtedly an example of a stereotype in the sense used by Labov (1994: 78). Glottal stops similar in form to those found in other dialects (i.e., fully voiceless and with a complete occlusion) are relatively rare. Docherty and Foulkes (1999, 2005) discuss subtle variation in production of glottal and glottalized forms in Tyneside English, but for the purposes of the present study we group together as a single category all types of variant which contain a laryngealized element.

Patterns of variant usage are constrained by a wide range of internal and external factors, including style (e.g., formality of speech), lexical identity (especially with respect to [ɹ] for /t/; Wells 1982), social characteristics of the speaker and interlocutor, conversational function (e.g., in turn endings—see also Local et al. 1986) and prosody (e.g., articulation rate and phonological context, Docherty 2007).

We focus here on two patterns which were especially clear in previous analyses. Both relate to correlations between variant usage and gender. First, in word-medial intersonorant contexts two main variants are found in the community: plain [p t k] and the local laryngealized variants. In free conversation the latter dominates for most speaker groups, while the plain variants are largely restricted to females (see Figure 14.1, which shows quantified data

drawn from the PVC project). As expected for a stereotype variable, the laryngealized forms decrease in more formal styles, being instead replaced by the standard plain variants. For older females in particular the style shift is absolute; that is, local laryngealized variants are not used at all in formal speech styles. Thus, notwithstanding style, age, and class effects, we can observe that plain variants are statistically much more likely to occur in female speech.

In word-final pre-pausal context a different set of variants is found. Here laryngealized or glottal forms are very rare. Instead, plain [p t k] are the default forms for most speakers (a pattern which, incidentally, differentiates Tyneside from most other British accents, where glottal stops are usually frequent in this context). For young women, though, pre-aspirated [ʰp ʰt ʰk] are emerging as a favored form. Our analysis of word-list data showed that 70% of tokens produced by young women were pre-aspirated (Docherty and Foulkes 1999). Pre-aspiration was considerably less frequent for younger males (35%). The overall rate of pre-aspiration was also much lower for older speakers in

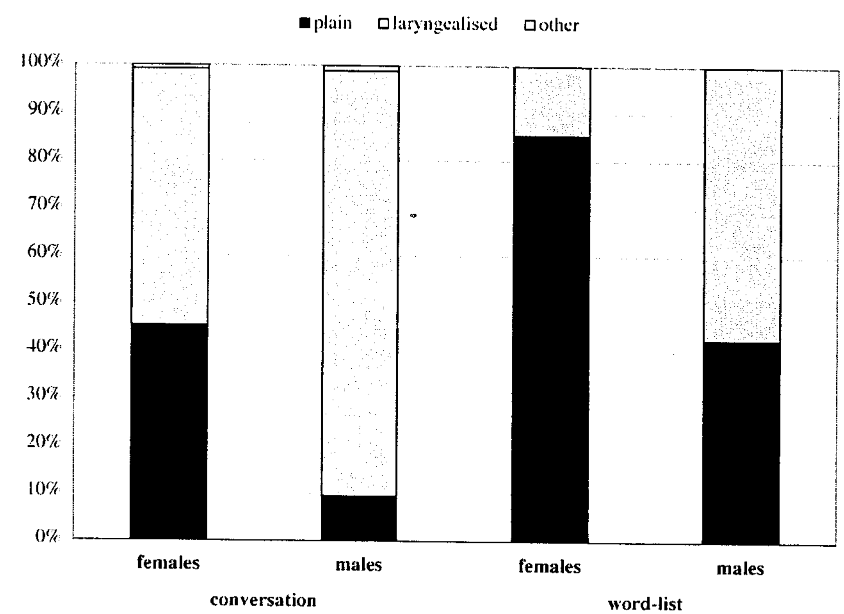

Figure 14.1 Variant usage for word-medial (p,t,k) by sex and speaking style, Tyneside adults (N tokens: conversation = 1,764 for females, 1,628 for males; word-list = 571 for females, 572 for males).

the PVC sample, but the sex pattern was consistent (23% for older women compared with less than 1% for the older men).

Variant usage is of course influenced by a complex range of factors, as noted in passing above, and any generalization about a single factor inevitably represents an oversimplification. However, the patterns reported here for Tyneside are sufficiently clear for us to infer that the variants of voiceless stops are indexical of gender. Plain variants in intersonorant position and pre-aspirated variants in pre-pausal context are far more frequent in the speech of females than males.

The question at issue in the remainder of this study is whether members of the speech community under investigation can recognize such patterns. That is, do native users of the Tyneside dialect infer patterns of indexicality from sociolinguistic variables akin to those which have been identified in studies of speech production? In our specific case example, do Tynesiders recognize that males and females use variants of voiceless stops in statistically different ways? Do they have implicit or explicit knowledge of patterns of variation, such that they can associate a particular variant with aspects of a speaker's social background?

4. Method

As noted earlier, the experiment was designed to probe the role of various phonetic cues in listeners' identification of a talker's sex. Assuming that f0 would not differ markedly for children, we constructed a simple listening test using samples of children's speech as stimuli. The task for listeners was to indicate whether they judged the child to be a boy or a girl.

4.1 Speakers

In our corpus of children's speech we identified six individuals who produced the full range of sociophonetic variants under investigation. That is, examples were found in each child's recording of both plain and laryngealized variants in word-medial context, and both plain and pre-aspirated variants in pre-pausal contexts. Three boys and three girls were chosen to provide a balanced sample. The children were among the older ones in the corpus, with an age range from 3;0 to 4;1, although the age of the child was not used as a criterion for selection. Instead we targeted those children whose speech was relatively well developed and fluent, and whose recordings contained ample material suitable for the experiment.

4.2 Stimuli

From the six children's recordings we extracted a total of 67 tokens for the listening test (Table 14.2). The extraction was performed using the editing tools in the Praat software program. The tokens extracted consisted mostly of single words, with four of the tokens being two word phrases. 27 tokens were single syllable words, 33 contained two syllables, five contained three, and two were four syllable items. Fifty-three of the tokens contained a voiceless stop in medial or final position, with the other 14 being fillers. We were not concerned with including the same number of tokens for each variant or context. The 53 tokens of interest were selected because they were the only ones available in the six recordings which (i) we could extract without problem (e.g., they were spoken in isolation rather than in a continuous sentence, which might lead to clipping of part of the extracted token), and (ii) which were not affected by background noise or overlapping speech from someone else present during the recording. Of the total of 67 tokens, 32 (47.7%) were taken from boys and 35 (52.3%) from girls.

Table 14.2 Stimuli Used in Listening Test

stimuli	variants	n	variants	n	N	examples
medial	plain [p t k]	20	laryngealized [b̰ d̰ g̰]	12	32	*jumper, letters, chicken, the water*
pre-pausal	plain [p t k]	7	pre-aspirated [ʰp ʰt ʰk]	14	21	*up, cat, look, fell out*
fillers					14	*bath, bumble bee, blue*
Total					67	

The extracted tokens were compiled into a new (.wav) sound file for the purposes of the listening test. Each stimulus in the test consisted of three consecutive repetitions of a token. A gap of 1.5 seconds was inserted between each token repetition, and a longer gap of 5 seconds separated adjacent stimuli.

4.3 Listeners

Three listener groups were recruited for the experiment. The listeners were drawn from staff and students at the home universities of the authors. Some students who participated had some training in linguistics. However, none of the students was at an advanced level and none knew either the purpose of the experiment or had awareness of the first two authors' previous work on Tyneside English. The first (experimental) group consisted of 20 natives of the Tyneside region. The other two groups were selected as controls for comparison with the Tynesiders. One consisted of 35 native British English speakers from regions other than Tyneside. They are referred to henceforth as the "non-local UK" listener group. Although we would predict some familiarity with varieties of British English, we assumed this group would have little awareness of the indexicality of sociolinguistic variables specific to Tyneside. The other control group consisted of 114 American students. These participants came from a range of geographical backgrounds but were all resident at the time of the experiment in Tucson, Arizona. We assumed that this group had little or no knowledge of phonological variation in British English. No formal testing was carried out of speech or hearing disorders, but none of the listeners reported any such problems.

4.4 Listening tests

The tests were conducted on campus at the participating universities. The British listeners participated in the tests in computer laboratories. The test sound file was played through standard audio programs, with listeners wearing good quality headphones. The American group heard the sound file via high quality amplification in a classroom setting. Although not wearing headphones, both laryngealization and pre-aspiration could be heard clearly by the administrator (the fourth author, who was positioned furthest away from the amplifier).

Listeners were given an answer sheet consisting of a transcription of the stimulus and two responses, "boy" and "girl" (Figure 14.2). The structure of the test was outlined verbally by the test administrators, who also explained that all the children came from Newcastle upon Tyne, UK. Listeners were instructed to judge whether the speaker of each stimulus was a boy or a girl, and to circle the appropriate answer. They were warned that they would find the test difficult, but further instructed to provide an answer for each stimulus even if they had to guess to do so.

1. cat	BOY	GIRL
2. letters	BOY	GIRL
3. bath	BOY	GIRL

Figure 14.2 Sample of answer sheet.

A training test was provided which consisted of three stimuli presented in the same format as the main test. After the completion of the training test listeners had the opportunity to ask questions about the test (and, in the case of the British groups, the audio programs). In practice the only questions raised concerned modification to the volume of sound playback. The main test was thus administered a few minutes after the training test. Despite reports that they found the task difficult, listeners performed as instructed, offering answers to each stimulus with only two exceptions.

4.5 Analysis

While our main point of interest was in the effect of the sociolinguistic variants on sex identification, we anticipated that the stimuli would also display variation along other parameters which might influence listener response. Following discussions with some of the participants after the test, and in light of predictions derived from previous linguistic research, we coded the stimuli for a number of factors. In addition to the sociolinguistic variant, each stimulus was also measured for f0 and amplitude. Quantification of f0 was performed for the obvious reason that f0 is a key cue to speaker sex for adults, and it was therefore possible that f0 differences might influence response for the child data. f0 was measured in Praat, recording the average f0 value across the whole stimulus. Amplitude was also measured in Praat and recorded as a mean for the stimulus, as a reflection of its overall loudness. Two further factors were coded for auditorily: articulation rate and voice quality. For speech rate the first three authors judged each token to be "normal," "slow," or "fast." For voice quality we recorded judgments of "modal," "breathy," and "creaky." In both cases the majority decision was taken as the final classification. We restricted ourselves to these simple taxonomies because the stimuli seemed to us to differ most clearly through phonation type, and because these labels approximated the kinds of comments listeners reported when discussing how they arrived at their responses. We subsequently made a more objective analysis of rate in terms of syllables per second. We made no attempt to analyze formant values or spacing, since previous studies which have identified formant differences in child speech have done so with fully controlled materials such as prolonged vowels spoken in isolation (Bennett and Weinberg 1979b, Perry et al. 2001).

Statistical analysis involved binary logistic regression in the first instance in order to explore the overall variance in the data. Regression analyses were carried out for each listener group, with separate runs for word-medial responses and pre-pausal responses. The dependent variable was response ("boy" or "girl"), and the independent variables were amplitude, f0, rate, voice quality and variant (plain versus laryngealized for the run on medial responses, plain versus pre-aspirated for pre-pausal responses). Post hoc analysis was performed using chi square tests.

5. Results

Since the listening test yielded a binary outcome we report all results arbitrarily in terms of "girl" responses.

5.1 Overall responses

Although we were not concerned with how accurate listeners were in identifying the sex of the speaker, it is worth reporting that the proportion of correct responses was very similar for the three listener groups (Table 14.3). Moreover, the figures approached chance level at 50%, which was perhaps to be expected. The proportion of "girl" responses, however, was slightly higher than "boy" responses for all three groups (recall also that more of the tokens did in fact come from girls' speech). Neither the correct responses nor the "girl" responses differed significantly across the two British listener groups. However, the American group gave significantly fewer correct responses than the non-local UK group (chi sq = 5.992, df = 1, $p < .025$).

Table 14.3 Overall Distribution of Results

listener group	correct responses (%)	"girl" responses (%)	N
Tynesiders	48.7	52.0	1,340
non-local UK	49.4	53.3	2,343
Americans	46.5	51.4	7,648

5.2 Logistic regression analysis

The results of the exploratory logistic regression analyses are summarized in Table 14.4. Unsurprisingly the results were complex, as is to be expected when we consider the number of factors included in the analysis, and the fact that the stimuli were both small in number and relatively uncontrolled. All factors were returned as significant in one or more of the runs for at least one of the listener groups.

Table 14.4 Factors Returned as Significant in Logistic Regression Analysis

word-medial stimuli					
group	**amplitude**	**f0**	**voice quality**	**rate**	**variant**
Tynesiders	**				
non-local UK	***	**			
Americans	***	***	***	(*)	
pre-pausal stimuli					
group	**amplitude**	**f0**	**voice quality**	**rate**	**variant**
Tynesiders	(*)	**	***	*	
non-local UK	***	***	***	**	
Americans	(*)	***	**	***	***

(*) $p < .1$, * $p < .05$, ** $p < .01$, *** $p < .001$

We refrain from reading too much into the regression results, since further tests would need to be carried out using more controlled data to clarify the patterns observed. What is clear, however, is that the sociolinguistic variants do play a role in listener response, albeit a relatively small one compared with other factors, and much more clearly for pre-pausal responses than for medial ones. In light of the overall complexity in the data we continue with an exploration of responses in relation to each of the main factors.

5.3 Results by amplitude

The regression analyses showed amplitude to be a significant factor in the responses by non-local UK listeners for both medial and pre-pausal stimuli, and by Americans for medial stimuli (Table 14.4). Figure 14.3 enables us to take a closer look at the patterns in the responses. It shows responses in the form of scatter plots, for the three listener groups separately. Results for medial and pre-pausal stimuli are pooled. The vertical axis represents the proportion of 'girl' responses given by listeners, while the horizontal axis indicates the mean amplitude of the stimulus (in dB). Each data point indicates responses to an individual stimulus. Trend lines are also included for each listener group.

Figure 14.3 indicates a consistent and clear pattern for all three groups. While relatively louder tokens are readily perceived as being either "boy" or "girl," relatively quieter tokens (those with lower amplitude values) elicited more "girl" responses. The effect is clearest for the quietest tokens, i.e., those to

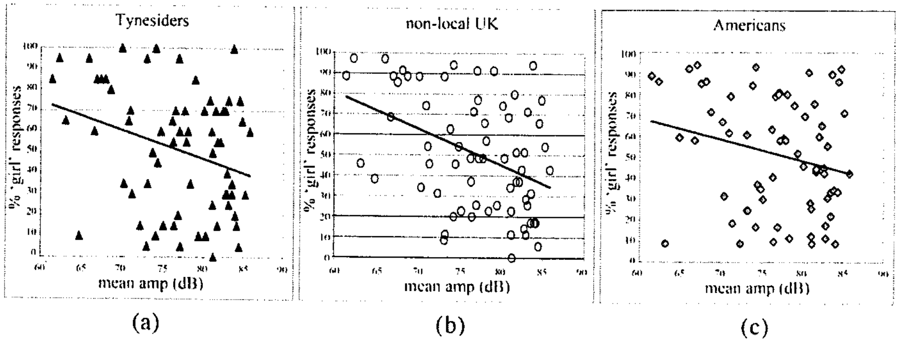

Figure 14.3 Percentage of "girl" responses by amplitude of stimulus; (a) Tynesiders, (b) non-local UK, (c) Americans.

the left hand side of the scatter plots. Results of correlation analyses are summarized in Table 14.5. The negative correlations are significant for all listener groups when all tokens are considered together. When medial tokens are examined alone the correlations are again significant for all three listener groups. For pre-pausal tokens taken alone the correlation is significant only for the non-local UK group, but the negative trend is maintained for the other groups.

Table 14.5 Correlation Analysis Results, Responses by Amplitude

		Tynesiders	non-local UK	Americans
all stimuli (df=65)	*r*	–.300	–.409	–.236
	p	< .01	< .0005	< .05
medial only (df=30)	*r*	–.362	–.434	–.310
	p	< .025	< .01	< .05
pre-pausal only (df=19)	*r*	–.273	–.466	–.137
	p	n.s.	< .025	n.s.

5.4 Results by f0

f0 was returned as a significant factor in five of the six regression analyses (Table 14.4). Closer analysis of the results, however, fails to reveal a clear pattern. Figure 14.4 shows responses in the form of scatter plots, following the same format as Figure 14.3.

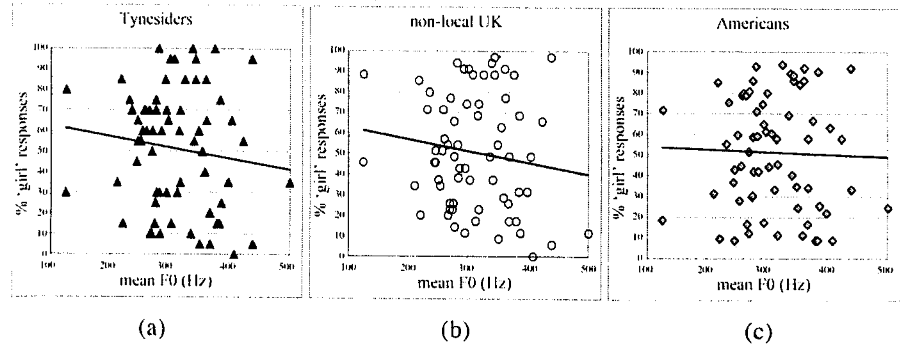

Figure 14.4 Percentage of "girl" responses by f0 of stimulus; (a) Tynesiders, (b) non-local UK, (c) Americans.

For adults high f0 is usually associated with female speakers and low f0 with males. In light of this we might predict that stimuli with high f0 would be more likely to elicit "girl" responses from listeners in our test. This was not the case in our test. The trend lines in fact indicate an effect in the opposite direction, with low f0 stimuli eliciting more "girl" responses from all listener groups. However, none of the correlations were close to significance. The plots in Figure 14.4 present all stimuli together, with no division according to the phonological context of the voiceless stop. Further exploration of the f0 data with reference to phonological context fails to clarify the picture. When stimuli with a word-medial or pre-pausal /p t k/ are analyzed separately the correlations remain nonsignificant.

It is clear that to understand the relationship between f0 and listener response requires further research. We might offer a partial explanation for our findings, however, with reference to the relationship between f0 and amplitude. As Figure 14.3 showed, quiet stimuli elicited more "girl" responses while more "boy" responses were given for louder tokens. As is well known, louder speech typically leads to an increase in f0, as the increased airflow required to raise amplitude will also (unless the speaker makes compensatory adjustments) lead to faster rate of vocal fold vibration. Quiet speech may by contrast involve relatively low f0. What we see in Figure 14.4, then, might in fact be an indirect reflection of the loudness of the stimuli. We therefore tested whether there was any correlation between the f0 and amplitude measures in our data. When all stimuli were considered there was indeed a positive—but non-significant—correlation. That is, louder stimuli had higher f0, and quieter stimuli lower f0, but the effect was not marked. We also considered the fact that listeners might only show a clear pattern of responses for stimuli with extremely high or low f0 values. When we examined the f0 and amplitude measures in more detail we found a much stronger positive correlation when only those stimuli with extremely low f0 values were considered. We arbitrarily tested for f0 and amplitude correlations with the 10, 12, 15, and 20 stimuli which had lowest f0. The correlation coefficient reached significance for the 12, 15, and 20 stimuli with lowest f0. No significant correlations were found in a similar set of comparisons with the highest f0 stimuli. Thus, in summary, stimuli with very low f0 may be eliciting high numbers of "girl" responses because they are also quiet. This remains, however, an issue which demands further exploration in future experiments.

5.5 Results by voice quality

Voice quality was not generally identified as a significant factor in the regression analyses, the only exception being for American listeners with pre-pausal stimuli (Table 14.4). Nevertheless we present the main effects here for completeness, and because a predictable trend did emerge in the responses. Figure 14.5 represents the proportion of "girl" responses for the three groups, with tokens divided according to whether they were coded as modal, creaky, or breathy. The number of stimuli in the creaky and breathy categories is small (6 creaky and 4 breathy from the total of 67), but it is noteworthy that the breathy tokens elicited the highest proportion of "girl" responses from all groups.

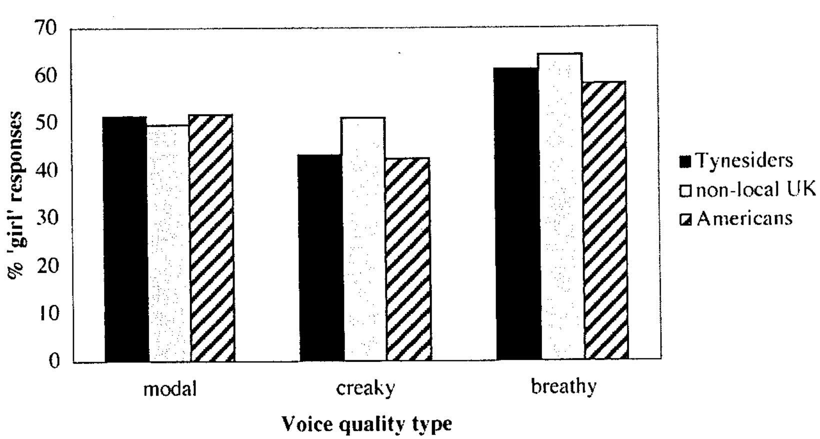

Figure 14.5 Percentage of "girl" responses by voice quality.

5.6 Results by articulation rate

The data are arranged in Figure 14.6 according to our auditory categorization by articulation rate. As was the case with voice quality, the number of stimuli in the non-modal categories is fairly small (n=13 in each case). The emergent trend in this particular analysis was variable across the listener groups. Both British groups gave fewer "girl" responses to fast stimuli than they did slow stimuli. For the Americans the reverse pattern was found.

Following the regression analysis we took an objective measure of articulation rate, in terms of syllables per second for each token (Künzel 1997). The results of this more detailed analysis failed to clarify the picture, however, with no significant correlation emerging between articulation rate and responses.

5.7 Results for word-medial variants

No variant was returned as a significant factor in the regression analysis for stimuli with medial /p t k/. However, when we compare the listener groups we do find evidence for a significant difference in responses.

Figure 14.7 shows the proportion of "girl" responses for medial tokens containing plain and laryngealized variants separately. For both the non-local UK

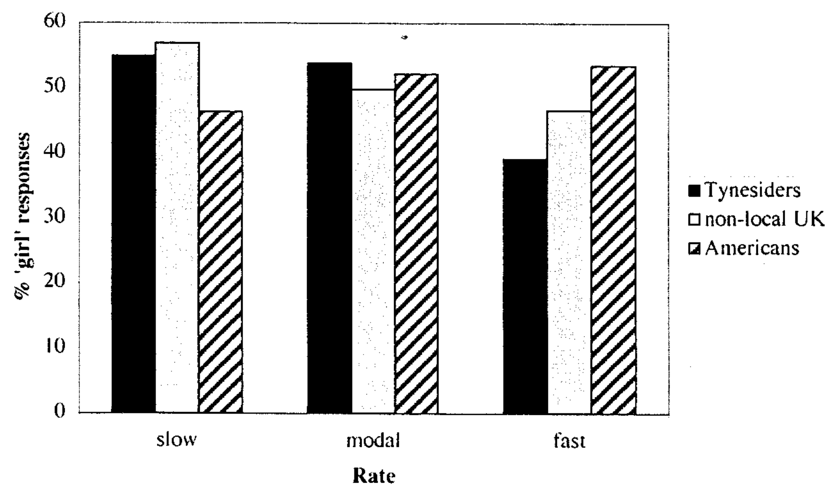

Figure 14.6 Percentage of "girl" responses by articulation rate.