The perception of personal identity in speech: Evidence from the perception of twins' speech

Keith Johnson
Misty Azara

Department of Linguistics
Ohio State University
222 Oxley Hall
1712 Neil Ave.
Columbus, OH 43210-1298

Running head: Perception of twins' speech

Received:

Abstract

Three experiments studied the perception of speech produced by female twins (5 MZ, 1 DZ) ranging in age from 20 to 67. The experiments were designed to discover whether listeners could detect differences between twins, and to explore listeners' perceptual representations of talkers. Because identical twins have virtually identical vocal tracts and the twins in this study grew up together in the same home, they serve as a unique control population for studies of the perception of personal identity. The results showed that listeners' sensitivity to twin differences was greater than chance and stable over changes in experimental conditions. Analysis of the perceptual space for talkers showed that the difference between identical twins was in some cases as large as the difference between unrelated talkers. The dimensions of the perceptual space were related to age and dialect, and the distance between twins in the perceptual space was not related to age.

PACS numbers: 43.71.B, 43.70.G

INTRODUCTION

It is commonly asserted that personal information in speech is determined by both anatomical and behavioral factors. For example, Ladefoged & Broadbent (1957: 98) asserted that "the idiosyncratic features of a person's speech" may "be a part of an individual's learned speech behavior" and may also "be due to anatomical and physiological consideratios." Table 1 lists some anatomical and behavioral sources of talker variation and indicates the key question addressed in this paper with the question mark next to ideolect. The strategy adopted in this paper is to ask whether ideolect is perceptible to listeners when the other factors in table 1 are held constant.

Insert Table 1 about here

Garvin & Ladefoged (1963) made explicit a tactical decision on talker variability that lies at the heart of most speech perception research conducted in the preceeding and subsequent years.

> "In this connection, it is important to note that the acoustic analysis of the primarily organic conditions seems to present many fewer difficulties than that the exclusively learned characteristics. Organic conditions may be amenable to analysis in terms of a few typical features, while learned characteristics undoubtedly will require an extremely detailed manipulation of a large variety of features. The study of a complex variable such as the differences in learned patterns will initially have to be avoided by the simple device of holding that particular condition constant." (p. 198)

One radical stance on talker variability would be to accept that Garvin & Ladefoged's tactic as reality - all talker variability is "organic". In this view, anatomical factors are taken to determine all individual differences among speakers of the same dialect (Nordström & Lindblom, 1975; Nearey, 1978; Syrdal & Gopal, 1986; Miller, 1989). Of course, behavioral differences distinguish speakers of different dialects, so that in one dialect community *coffee* is pronounced [kAfi] while in another it is pronounced [kwOfi]. But in the radical invariance view, all individual acoustic

differences among speakers of the [kwOfi] dialect, for example, are solely due to differences in vocal tract anatomy. The assumption here is that speakers of a dialect differ only in terms of anatomy and not in articulatory behavior. Thus, perceived talker identity is determined by factors such as voice pitch and mode (breathiness, etc.), and by vocal tract resonant frequencies reflecting the overall length of the vocal tract. An approach along these lines is at the heart of the radical invariance theory of vowel normalization (see Johnson, Strand & D'Imperio, in press). In this view of vowel perception, a normalization procedure removes the effects of anatomical differences between talkers and the resulting vowel representations are assumed to be invariant from speaker-to-speaker.

A somewhat more subtle deterministic account that allows for some variation in articulatory behavior is also possible. It could be that vocal tract anatomy indirectly determines speech production patterns. This indirect causation account is no less deterministic than the radical invariance view, but predicts lawful articulatory variation across speakers. Consider for example Johnson, Ladefoged and Lindau's (1993) discussion of individual differences in an X-ray microbeam study. They found that, in a dialectically homogeneous group of speakers, some showed a large jaw height component distinguishing high and low vowels while others were not 'jaw-movers'. Johnson, Ladefoged and Lindau considered the possibility that the jaw-movers had less deeply domed palates than the jaw-nonmovers. This is an example of an indirect causation account of talker variability. In this account, vocal tract geometry determines the particular patterns of interarticulator coordination which will be most efficient or effective for a talker, and thus the dimensions of the vocal tract indirectly determine the different articulatory movement patterns observed across talkers.

The perception of talker identity in the indirect causation account involves somewhat more detailed vocal tract perception than envisioned in the radical invariance view. Listeners have available in the acoustic signal not only evidence concerning the overall length of the vocal tract and vocal folds, but also evidence (in gestural coordination patterns) concerning finer-grained details of vocal tract geometry. In this account of talker perception the vocal tract is the distal object of talker perception cued by complex, temporally distributed acoustic properties. Some of these properties are a reflection of gross anatomical features such as vocal tract length while others are behavioral variations in gestural coordination. Note, however, that in this account such behavioral variation is

still determined by vocal tract geometry - vocal tract geometry is merely signaled indirectly.

Contrasting with these two deterministic accounts of individual differences in speech production is a view that talker variability is to a certain extent idiosyncratic, reflecting private motivations for behavioral variation (c.f. Hocket's, 1958, discussion of 'ideolect'). This is the definition of "personal information" assumed by Ladefoged & Broadbent (11958) in the quotation at the start of this paper. Some evidence from studies of speech articulation support this view. Examples include tongue-tip up or down /s/ production (Borden & Gay, 1979), degree of jaw recruitment in low vowel production (Johnson, Ladefoged & Lindau, 1993), bunched vs. retroflex /r/ (Hagiwara, 1995), and voice onset time (VOT) duration (Lisker & Abramson, 1964; Newman, 1996). These individual differences in speech production are apparently not determined by anatomy or dialect, but rather reflect the speaker's individual speaking strategy. Such private strategies, if they exist, may arise during phonetic acquisition: one talker may move the jaw more or less than another (or use a tongue-tip up or down /s/, etc.) because he/she discovered during phonetic acquisition that this articulatory strategy could be made to work for speech communication.

Talker perception in this view involves the perception of a talker's personal anatomical characteristics as in the radical invariance and indirect causation hypotheses. However, in addition to this, the perceived identity of the talker is also cued by behavioral characteristics reflective of ideolect. Clearly, idiosyncratic habits of speech must be constrained if speech communication is to be successful - language, in the normal sense of the term, cannot be private. But the speech perception system has been found to be robust over quite wide variations in dialect, synthesized speech (Greenspan et al., 198x), and rather extreme signal manipulations (Remez et al. 1981; Shannon et al., 1995). This robustness suggests that there may be considerable latitude for individual variation. Which is to say that the idiosyncratic talker variability hypothesis is not implausible.

Each of these three perspectives - radical invariance, indirect causation, and idiosyncratic variation - is possible because it is not currently known how much of talker variability should be attributed to anatomical factors and how much should be attributed to behavioral factors. The work reported here addressed this issue in three experiments on the perception of speech produced by twins. Two questions were addressed in the experiments. The first question was: Can listeners tell

twins apart by their speech? This question serves as a precursor for more detailed acoustic and articulatory study of twin speech. The second question was: Are perceptual differences between twins consistent across experimental manipulation of instructions and list composition?

Twin Speech

The role of ideolect in perceived talker identity can be assessed in twin speech because identical twins have virtually identical anatomy.

In an extensive study of vocal tract geometry in twins, Lundstrom (1948) concluded that anatomical differences between identical twins are of the same magnitude as left-right asymmetries within individuals. This close anatomical similarity of identical twins is reflected in similarities in twins' speech in terms of long term average spectrum (Alpert, et al., 1963), and fundamental frequency (Gedda, et al., 1960), infant cries (Ostwald, et al., 1962), and possibly articulatory patterns including speech disorders (Matheny & Bruggemann, 1972; Locke & Mather, 1989). In addition to anatomical similarities, the early language experience of twins raised together is similar (see Stafford, 1987). Consequently, a study of talker perception in speech produced by identical twins provides a natural control over anatomical sources of talker variation.

Of the three perspectives discussed above - radical invariance, indirect causation, and idiosyncratic variation - only idiosyncratic variation predicts that twins will be perceptually distinguishable. The radical invariance and indirect causation views predict that differences between identical twins raised together will be vanishingly small. One recent acoustic study of speech produced by twins provides support for the idiosyncratic variation perspective. Nolan & Oh (1996) found that twins differ in their pronunciations of /l/ in English.

The structure of the paper is as follows: Sections I-III describe three experiments on the perception of personal identity in twins' speech. In experiment 1 listeners discriminated pairs of isolated words produced by the same person, or by twins. In experiment 2 the talker-discrimination trials included tokens produced by the same person, twin comparisons, and comparisons of unrelated people. Experiment 3 differed from experiment 2 only in the instructions given to listeners. In experiments 1 and 2 listeners were told that the discrimination pairs could include twin comparisons, in experiment 3 listeners were not given this warning. Section IV reports a

comparison of perceptual sensitivity and bias in the three experiments, and section V reports a multidimensional scaling analysis of perceived talker identity in experiment 3.

I. EXPERIMENT 1

Six pairs of female twins (5 monozygotic, 1 dizygotic) read a set of isolated words and then listeners were asked to judge pairs of words, saying whether the words were produced by the same person or by twins. The paired comparisons were composed of (1) two repetitions of the same word by the same person, or (2) repetions of the same word by twins.

One rationale for this experiment is as a pretest for a phonetic study of twin's speech. If listeners are able to detect differences between twins then we are encouraged to look more closely at acoustic and articulatory records to determine the ways in which the twins' speech differs, but if listeners are unable to detect differences between twins we wouldn't expect to find any meaningful differences in acoustic or articulatory studies.

More importantly, this experiment is a test of the three hypotheses outlined in the introduction. Radical invariance and indirect causation predict that talker identity (within dialect) is determined by vocal tract geometry. If we find that listeners can detect a difference between twins then we have some support for the hypothesis that at least some individual differences in speech production are idiosyncratic.

The speakers recruited for this study represented a fairly wide range of ages and dialects. This variation among the speakers leads to two secondary predictions. First, we might predict that older twins will differ more from each other than younger twins, assuming that older twins' have had a wider divergence of linguistic experience. Second, we might also predict that the range of dialect variation across twin pairs might enhance the apparent similarity of twins' voices (which makes the experiment a stronger test of the indiosyncratic variation hypothesis). These predictions will be discussed in section V in connection with a multidimensional scaling analysis of talker discrimination data from experiment 3.

A. Method

1.Speakers.

Six pairs of female twins served as speakers in this and the following experiments. Their ages at the time of the recordings ranged from 20 to 67 years. Though we had intended to limit the

study to monozygotic twins, one pair (QB & CK) were determined to be dizygotic on the Nichols & Bilbro (1966) test for twin zygoticity. The other speakers were also given the Nichols & Bilbro test and were found to be monozygotic. The dizygotic twins were no more or less perceptually confusable than the monozygotic twins in the experiments reported here.

Table 2 shows further information about the speakers. The oldest speakers (BB & BP) lived quite different lives - BB stayed in Ohio her entire life, while BP lived with her husband (who was in the Army) in Okinawa and Florida. At the time of this recording BB & BP were living together and both worked in a large department store in Columbus. BB balked at first when asked to read for this experiment saying "you don't want us because we can't say *r* when it is the third letter of the word" (see Lewis & Thompson, 1992 and Locke and Mather, 1989, on the concordance of articulatory patterns in twin self-reports).

Insert Table 2 about here

---

2. Word list

Each of these speakers was asked to read the word list shown in table 3. The list was designed to illustrate a range of phonetic contrasts in English which have been found to show individual differences. Words with vowels and diphthongs following /h/ or /ɹ/ tested for individual differences in vowel production (Johnson, Ladefoged & Lindau, 1993). Words with post-vocalic /r/ tested for differences in /r/ which have been noted before (Hagiwara, 1995). Finally, individual differences in voice onset time (VOT) were tested with the words beginning with /p/ and /b/ (Newman, 1996).

Insert Table 3 about here

---

3. Recording

The words in table 3 were presented to the speakers in five different random orders as lists printed on five different sheets. Filler words were added to the beginning and end of each sheet to avoid initial and final intonation on the test words. The word-list readings were recorded in a quiet

room using a Shure SM10a head-mounted microphone and a Marantz PMD 222 portable cassette recorder.

4. Listeners

Ten undergraduate students at Ohio State served as listeners in the experiment. None of the listeners reported any history of speech or hearing problems and all were native speakers of American English. The listeners received partial course credit for their participation.

5. Procedure

A subset of the word list (shown in boldface in table 3) was selected for presentation to listeners. The selected words sampled a range of the phonetic contrasts in the list. The number of words tested in the perception experiment was limited to this subset in order to keep the duration of the experiment to a single one-hour session for the listeners in experiments 2 and 3.

For each of the 14 test words we selected 24 paired comparisons. Twelve of the pairs contrasted two repetitions of the word by the same talker. We chose productions from the third and fourth readings of the word list for these comparisons. The remaining twelve comparisons were of twins (for example, speaker QB paired with speaker CK). The items for these comparisons came from the third reading of the word list. The total number of trials was thus 336 (24 comparisons for each of the 14 test words).

The test words were digitized onto a PC at 22.05 kHz with 16 bit resolution and saved as sound files for on-line presentation to listeners.

Listeners were seated in a sound-attenuated booth and heard stimuli over headphones at a comfortable listening level. They entered their responses by clicking a mouse button on a computer screen. The stimuli were presented in pairs with a 3 second response-to-stimulus interval and the listener's task was to indicate whether the pair of stimuli in a trial was produced by the same person or by twins. Listeners were told that half of the pairs that they would hear contrasted words spoken by identical twins and half of the trials were of the same person repeating a word twice. Listeners were instructed to choose the "different talkers" response button for twin comparisons were warned that twins can sound remarkably similar to each other.

B. Results

Listeners' ability to detect the difference between twins was measured by taking d' values by listeners and by words using the tables published by Kaplan, MacMillan & Creelman (1978) for AX presentation. In these analyses, the false alarm rate is the probability that a listener will respond "different talker" to comparisons of the same talker repeating a word. Note that the 'signal' being detected in this analysis is the identity of the talker because in none of the trials was the same acoustic signal presented twice. So a correct 'different' response had to be based on detection of a talker difference.

For twin comparisons the average percent "different" responses was 62% and for same talker pairs the average percent "different" responses was 11%. The resulting average d' value was reliably different from 0 (chance performance) in t-tests by listeners and by words [average d' = 2.76, $t1(9) = 32.4$, $t2(14) = 26.2$, both $p < 0.01$].

C. Discussion

This experiment found that listeners detected the difference between twins at a level reliably above chance performance. This result is important because it documents that listeners can detect talker differences which are arguably not due to differences in vocal tract geometry or formative language experience. This finding is consistent with the hypothesis that there are individual differences in speech production due to idiosyncratic habits of pronunciation in addition to differences due to vocal tract anatomy.

However, the listeners' sensitivity to the twin comparisons in this experiment may have been enhanced by the experimental design. Listeners knew to expect twin comparisons and these were the only "different" pairs in the experiment. This could have led listeners to attend carefully to small talker differences that they might otherwise not notice. Also, based on these results the size of the perceptual difference between twins cannot be compaired to the perceptual difference between unrelated individuals. Experiments 2 and 3 address these concerns by including comparisons across twin pairs, where the paired speakers were unrelated to each other (experiment 2) and by not telling listeners to expect twin comparisons (experiment 3).

II. EXPERIMENT 2

This experiment was identical to experiment 1 in all respects except one. In addition to pairs of stimuli which contrasted twins with each other and pairs of stimuli produced by the same talker, we included pairs of stimuli which were repetitions of the same word by two unrelated female talkers (unrelated twins).

A. Method

1. Listeners

Nine undergraduate students at Ohio State served as listeners in the experiment. None of the listeners reported any history of speech or hearing problems and all were native speakers of American English. The listeners received partial course credit for their participation.

2. Procedure

In addition to the 24 paired comparisons (12 same talker, and 12 twin comparisons) used in experiment 1, this experiment included 12 additional comparisons of unrelated talkers (for example, speaker QB paired with speaker BB). For each word in the test list 120 such comparisons of unrelated talkers were possible. We chose 12 of these comparisons randomly for each word. As before, the tokens for these comparisons were taken from the third reading of the word list. The total number of trials was thus 504 (36 comparisons for each of the 14 test words).

B. Results

For unrelated talker comparisons the average percent “different” responses was 88%, for twin comparisons the average percent “different” responses was 58%, and for same talker pairs the average percent “different” responses was 10%.

As in experiment 1, the detectability of talker differences for twin comparisons and for unrelated talker comparisons was measured by taking d’ values by listeners and by words.

Listeners showed better than chance sensitivity to talker differences in pairs of words produced by unrelated talkers [average d’ = 4.46, $t1(8) = 32.8$, $t2(13) = 33.5$, both $p < 0.01$] and they also showed better than chance sensitivity to pairs of words produced by twins [average d’ = 2.77, $t1(8) = 31.8$, $t2(13) = 27.3$, both $p < 0.01$]. The average d’ values for twin pairs and unrelated pairs were also reliably different; listeners showed greater sensitivity to comparisons involving

unrelated talkers than to comparisons involving twins [$t1(16) = 10.5$, $t2(26) = 9.42$, both $p < 0.01$].

The d' value for twin comparisons (averaged over listeners) was about the same in this experiment (2.77) as in experiment 1 (2.76) and the variability of the d' estimates was similar. The average standard deviation was 0.34 in this experiment and 0.37 for experiment 1 (averaged over the items and subjects analyses).

C. Discussion

Like experiment 1, this experiment found that listeners could detect the difference between twins. The presence of trials comparing unrelated talkers had no effect on listeners' sensitivity to differences between twins. Also, as expected, the twins were more confusable than were unrelated talkers. The magnitude of the perceptual difference between unrelated pairs of talkers is about twice that of the perceptual difference between the pairs of twins. This suggests that the factors that distinguish unrelated talkers (anatomical and behavioral differences) are on average more potent perceptually than are the factors that distinguish twins (behavioral differences). We will return to this in section V.

Experiment 3 explored the possibility that sensitivity to twin differences was enhanced in experiments 1 & 2 because listeners were told to expect trials comparing twins.

III. EXPERIMENT 3

In this experiment, the listeners were not told that some trials would have twins paired with each other. In all other respects the experiment was identical to experiment 2. This experiment is a fairly stringent test of the detectability of twin differences because listeners had no *a priori* reason to believe that two very similar sounding productions of a word were not simply two productions by the same person.

A. Method

1.Listeners

Ten undergraduate students at Ohio State served as listeners in the experiment. None of the listeners reported any history of speech or hearing problems and all were native speakers of American English. The listeners received partial course credit for their participation.

2.Procedure

Unlike those in experiments 1 & 2, listeners in this experiment were not told that comparisons of twins would be included in the experiment. In all other details the method was identical to that of experiment 2.

B. Results and Discussion

For unrelated talker comparisons the average percent "different" responses was 83%, for twin comparisons the average percent "different" responses was 41%, and for same talker pairs the average percent "different" responses was 5%.

The d' values in this experiment were almost identical to those of experiment 2. Listeners showed better than chance sensitivity to talker differences in pairs of words produced by unrelated talkers [average d' = 4.51, $t1(9) = 22.7$, $t2(13) = 34.3$, both $p < 0.01$] and they also showed better than chance sensitivity for pairs of words produced by twins [average d' = 2.62, $t1(9) = 13.2$, $t2(13) = 19.0$, both $p < 0.01$]. The average d' values for twins and unrelated talkers were also reliably different; listeners showed greater sensitivity to comparisons involving unrelated talkers than to comparisons involving twins [$t1(18) = 6.7$, $t2(26) = 9.4$, both $p < 0.01$].

The listeners in experiment 3, as in experiments 1 and 2, were sensitive to talker differences between twins. This result is more striking than the results of the first two experiments because we did not tell the listeners that twin comparisons would be included in the experiment.

IV. SENSITIVITY AND BIAS

The average d' value for twin comparisons in experiment 3 was a little lower than the d' values for twin comparisons in the first two experiments (2.62 versus 2.77 and 2.76 in experiments 1 and 2). To determine whether this difference was reliable we performed analyses of variance (by listeners and by items) of the d' values for twin comparisons in the three experiments using a between-subjects design. This analysis found no reliable difference between the d' values as a function of experiment [$F1(2,26) = 0.38$; $F2(2,40) = 0.54$, both $p > 0.5$]. This indicates that sensitivity to the differences between twins was unchanged in the three experiments. To test whether listeners' response bias was affected by the different manipulations in experiments 1-3, β

(Green & Swets, 1966) was calculated for twin comparisons by items and by listeners in the three experiments and these bias estimates were submitted to between-subjects analyses of variance. These tests did show an effect of the experimental design on listeners' bias [$F1(2,26) = 3.29$, $p = 0.05$; $F2(2,40) = 8.3$, $p < 0.01$]. Post-hoc tests (Fisher's least-significant-difference) found that bias values in experiments 1 & 2 were no different from each other, while bias in experiment 3 was reliably lower [$p < 0.01$ by items, and $p < 0.05$ by subjects]. The average β and d' values from experiments 1-3 are shown in figure 1. This analysis indicates that bias was altered by a change in instructions (experiment 2 versus experiment 3), but not by a change in the composition of the set of stimuli (experiment 1 versus experiment 2); while sensitivity to twin comparisons was constant across the experiments.

Insert figure 1 about here

At first glance it seems completely unremarkable that sensitivity did not change as a function of experimental manipulations while bias did. This is what we would expect given the behavior of bias and sensitivity measures in psychoacoustic tests with similar experimental manipulations. However, the results did not have to be this way. Signal detection analysis was used to analyze data in which the 'signal' was not any particular acoustic property of the stimulus but rather a more abstract cognitive percept - the perceived identity of the talker. All of the pairs of stimuli presented to listeners were acoustically different from each other in a variety of ways. Listeners had to judge, presumably, the magnitudes and natures of these acoustic differences in order to determine whether the two stimuli in a trial had been produced by a single talker or by two different talkers. In this situation it might be expected that listeners' criteria for this judgment (and hence their sensitivity to the differences between twins) could have changed as a function of the composition of the stimulus list or their prior knowledge that twins were included in the comparisons. One indication that talker identity judgments are quite different from simple psychoacoustic judgments is that the instruction manipulation (experiment 2 versus experiment 3) produced a large change in β while a large change in response probability (from 1/2 'same' trials to

1/3 'same' trials, experiment 1 versus experiment 2) had no effect on β. In ordinary psychoacoustic experiments both of these manipulations affect β (Green & Swets, 1966).

The fact that sensitivity to twin contrasts showed no change over these manipulations suggests that the perception of talker identity is not open to manipulation - only response bias was influenced by the experimental manipulations we tested (see Mullennix, Johnson, Topcu-Durgun & Farnsworth, 1995). These analyses of sensitivity and bias found that perceptual talker information is stable over manipulations of response bias. This finding is consistent with a 'frame of reference' approach to speaker normalization (Johnson, 1989) where it is assumed that listeners normalize speech based on the perceived identity of the talker.

V. MULTIDIMENSIONAL SCALING

In experiments 2 and 3, all possible pairs of talkers were presented to listeners and each pair was judged as the "same talker" or "different talkers". The proportion of "same talker" judgments is thus an estimate of the perceived similarity of the two talkers. The proportion of "same talker" judgments from experiment 3 (where β was lowest) were entered in a multidimensional scaling (MDS) analysis to construct a perceptual map of the female talkers in this study. In this monotonic MDS analysis the Minkowski constant was 2 (i.e. we used Euclidian distances) and the Kruskal loss function was used. Figure 2 shows the two dimensional MDS solution. This solution accounted for 92% of the variance and the dimensions are easily interpretable, whereas the one dimensional solution, though it accounted for 83% of the variance, was not interpretable. The high proportion of variance accounted for indicates that the similarity measurements were internally consistent with each other despite the fact that they were based on different numbers of presentations (recall that in experiments 2 & 3 a random sample of the possible 'unrelated talker' comparisons was chosen for each test word).

Insert Figure 2 about here

Dimension 1 in the MDS solution is related to the age of the talker - the oldest speakers are

on the left side of the map while the youngest speakers are on the right side of the map in figure 2. There is almost a perfect rank order correlation between age and a diagonal from the upper right to the lower left of the perceptual map; only the two youngest pairs of twins (at ages 20 and 21) are out of order. Dimension 2 seems to be related to dialect: separating the two pairs of speakers from Oklahoma in the lower right quadrant of the map.

Are age and dialect the dimensions that listeners used to distinguish talkers from each other in this study, or would it be more accurate to find acoustic properties that correlate with the results of the MDS analysis? We expect that acoustic correlates could be found. For example, aspects of phonation such as breathiness might correlate with the first dimension for these speakers, and the movement of F2 during diphthongs (e.g. [aI] versus [A] in *hide*, for example) may be correlated with the second dimension. But it seems likely that other acoustic correlates of age and dialect would emerge from a replication of this study with a different set of speakers or words. That is, we predict that acoustic properties will correlate with the dimensions in a perceived talker space in just those cases where the acoustic properties are correlates of personal characteristics (such as age and dialect) that are important to listeners.

Two other patterns that might have been expected did not emerge in these data. First, we might have expected that older twins would be more easilyt distinguished from each other than younger twins because of divergent linguistic experience during adulthood. The perceptual data do not show this effect. The twins who were the most similar to each other were the oldest ones (BB & BP, age 67) and the youngest ones (AJ & NJ, age 20). The twins who were the most different from each other were the next oldest (RJ & AM, age 43) and next youngest (MN & CN, age 21). It is interesting that there was no tendency for twins to differ increasingly from each other with increasing age. If there had been such a tendency it could have been attributed to divergent linguistic experience during adulthood. Should the lack of age-related divergence be considered evidence that linguistic experience during adulthood has little impact on a person's habits of articulation? The results of this analysis are consistent with this interpretation. Second, the distance between the fraternal twins in this study (CK & QB) was not larger than the typical distance between the identical twins. This is interesting because fraternal twins have more divergent vocal tract anatomy than do identical twins (Lundstrom, 1946). This suggests that the difference between

identical twins is at least as large as what we would expect from certain vocal tract anatomy differences.

This leads to the question of idiosyncratic variation in speech production. The results of experiments 1-3 have demonstrated that listeners can detect differences between twins at a level better than chance, though listeners are more likely to confuse twins than they are to confuse unrelated speakers. Building on this finding, the perceptual map in figure 2 provides one estimate of the magnitude of idiosyncratic variation. Consider, for example, the locations of speakers LW and NJ in figure 2. These two speakers were closer to each other than any other non-twin pair of speakers. LW and NJ were both closest to their twins in the perceptual map, but the distance between them is smaller than the distance between two of the pairs of twins (RJ & AM and MN & CN). This would lead us to conclude that sometimes the distance between twins, who share vocal tract anatomy and formative language experience, is as large as the distance between unrelated individuals.

VI. General Discussion

Experiments 1-3 found that listeners can distinguish isolated words produced by twins at better than chance. Sensitivity to twin differences was unchanged by manipulation of stimulus list or instructions. Bias was also unchanged by stimulus list composition, but was changed by instructions. A multidimensional scaling analysis of perceived talker similarity showed that sometimes twins are perceived to be as different from each other as are unrelated talkers.

The experiments reported in this paper were designed to test the hypothesis that individual differences in speech production are due, at least in part, to idiosyncratic habits of pronunciation, in addition to variation among talkers due to differences of vocal tract geometry or differences in language or dialect. The radical invariance and indirect causation hypotheses attribute within dialect talker variability exclusively to anatomical factors, while the idiosyncratic variation hypothesis allows for ideolectal habits of pronuncation. The results of the present experiments support the idiosyncratic variation hypothesis. What is more, these results suggest that idiosyncratic talker variation can be quite large. Perceived talker differences between twins can be as large as those due to vocal tract and dialect variation.

These findings are important for both theory and practice in several areas of phonetic

research. These implications are discussed in the following sections.

Talker identity.

In gender studies it is commonly held that people "perform gender", that personal identity as male or female is to a certain extent the result of performance (Strand, 1999). The present findings are consistent with this perspective, suggesting that talkers have and use wide latitude in conveying personal identity. The results show that conveyed personal identity is determined both by anatomical features of the vocal aparatus (twins tended to be perceptually similar to each other) and by idiosyncratic factors (twins differed from each other). This is not to say that the performance of personal identity is entirely conscious, though conscious intent may certainly be involved. But the magnitude of idiosyncratic variation in this study was surprizingly large.

At a more practical level in the study of personal identity in speech, the experiments reported here exemplified one approach to probing listeners' perceptual dimensions for talkers' voices. Our approach built on Krieman & Gerratt's (1996) multidimensional scaling studies of the perceptual representation of pathological voices. The fact that we, like Krieman & Gerratt, find an orderly interpretable perceptual space suggests that MDS is a promising technique for the study of talker identity. The orderliness of the perceptual space for talkers also suggests that similarity of a new talker to an old talker may provide listeners with a way to establish a perceptual frame of reference for the new talker (Ladefoged & Broadbent, 1958; Johnson, 1990).

The MDS method could also be used in studies of the role of gender stereotypes in speech perception (Strand, 1999). It may be that gender stereotypes warp the perceived talker space in ways that can not be predicted by acoustic properties alone. In this connection for example, it is well known that perceived talker identity is inaccurate at least on dimensions such as the height and weight of the talker (Lass, 19xx).

Talker normalization in speech perception.

Theories of speech perception also have to consider individual differences in speech production because listeners recognize words despite wide variation in the acoustic signals produced by different talkers. The degree to which this variation is lawfully related to vocal tract geometry determines to a large extent the type of theory one can adopt. If variation among talkers is

idiosyncratic, then theories that assume a kind of "vocal tract" normalization (i.e. that the listener calibrates speech produced by different talkers based on perceptual representations of their vocal tracts) can account for only a part of what the listener accomplishes in speech perception.

Vocal tract normalization, the dominant approach to the speaker normalization problem, rests on the assumption that idiosyncratic talker differences are of minor importance. The present study tested this assumption and found that talker variability is much more than variation in vocal tract geometry and early language experience. This finding presents a potentially major challenge to the vocal tract normalization approach. The full scope of the challenge will be more apparent when the acoustic/phonetic nature of twin speech differences are more carefully studied.

Speech recognition.

Research in speech technology (particularly in spoken word recognition) also must consider the range and nature of individual differences in speech production. Talker variability is one of the chief practical hurdles in the development of robust speech recognizers, and much effort has gone into developing systems that adapt to talkers' pronunciation patterns. One avenue of research in this area has been to classify talkers by group membership and tune the recognizer based on this classification. If the magnitude of idiosyncratic talker variation is very large this approach may be unacceptable.

Further research.

The results reported here call for further research. First, perceptual results were used to study individual differences in speech production. This approach is justified because listeners are sensitive to a wider range of information than is usually measured in acoustic or articulatory studies. Now that we know that twins can be distinguished by listeners, the next step is to study patterns of speech production in twin speech. Second, forensic phonetic studies of talker identification (Hollien, 1990) have found that relatively long stretches of speech (several seconds) are necessary for accurate talker identification. The perceptual map in section V might have been different if longer utterances, instead of isolated word stimuli, had been used. Finally, the experiments used a non-homogeneous group of talkers. Even though this tends to enhance non-twin comparisons relative to twin comparison, it may be the case that a more homogeneous group of

talkers would be a more meaningful test of twin versus non-twin differences.

ACKNOWLEDGEMENTS

This research was supported by the Department of Linguistics and the Office of Research at Ohio State University; and by the National Institute of Deafness and Communication Disorders, Grant No. 7 R29 DC01645. We thank our colleagues in the departments of Linguistics, Speech and Hearing Science, and Psychology at Ohio State for valuable comments on this research, particularly Mary Beckman, Mark Pitt, and Ilse Lehiste.

References

Alpert, M., Kurtzberg, R.L., Pilot, M. & Friedhoff, A.J. (**1963**). "Spectral characteristics of the voices of twins," *Acta Genetica Med. Gemellol.* **12**, 335-341.

Borden, G.J. & Gay, T. (**1979**). "Temporal aspects of articulatory movements for /s/-stop clusters," *Phonetica* **36**, 21-31.

Garvin, P.L. & Ladefoged, P. (1963) Speaker identification and message identification in speech recognition. *Phonetica* **9**, 193-199

Gedda, L., Fiori-Ratti, L. and Bruno, G. (**1960**). "La voix chez les jumeauz monozygotiques," *Folia Phoniatrica* **12**, 81-94.

Green, D. & Swets, (**1966**) *Signal Detection Theory*.

Greenspan, S., Nusbaum, H. & Pisoni, D.B. (**19xx**) Perceptual learning of synthetic speech. JEP:HPP

Hagiwara, R. (**1995**). "Acoustic realizations of American /r/ as produced by women and men," UCLA Working Papers in Phonetics **90**.

Hockett, C.F. (**1955**). *A Manual of Phonology*. (Memoir 11, Indiana University Publications in Anthropology and Linguistics, Bloomington, IN).

Hollien, H. (**1990**). *The Acoustics of Crime: The New Science of Forensic Phonetics*. (Plenum Press, NY).

Johnson, K. (**1990**) The role of perceived speaker identity in F0 normalization of vowels. *J. Acoust. Soc. of Am.* **88**, 642-654.

Johnson, K., Ladefoged, P. & Lindau, M. (**1993**). "Individual differences in vowel production," *J. Acoust. Soc. of Am.* **94**, 701-714.

Kaplan, H.L., MacMillan, N.A. & Creelman, C.D. (**1978**). "Tables of d' for variable-standard discrimination paradigms," Behavior Research Methods & Instrumentation **10**, 796-813.

Kreiman, J. & Gerratt, B.R. (**1996**). "The perceptual structure of pathologic voice quality," *J. Acoust. Soc. of Am.* **100**, 1787-95.

Ladefoged, P. & Broadbent, D.E. (**1958**) Information conveyed by vowels. *J. Acoust. Soc. Am.* **29**, 98-104.

Lewis, B.A. and Thompson, L.A. (**1992**). "A study of developmental speech and language disorders in twins," J. of Speech and Hearing Res. **35**, 1086-94.

Lisker, L. & Abramson, A.D. (**1964**) A cross-language study of voicing in initial stops: Acoustic measurements. *Word* **20**, 384-422.

Locke, John L. and Mather, Patricia, L. (**1989**). "Genetic factors in the ontogeny of spoken language: Evidence from monozygotic and dizygotic twins," Journal of Child Language **16**, 553-9.

Lundstrom, Anders (**1948**). *Tooth Size and Occlusion in Twins*. (S. Karger,Basle).

Matheny, A. P., Jr., & Bruggemann, C.E. (**1972**). "Articulation proficiency in twins and singletons from families of twins," *Journal of Speech and Hearing Research* **15**, 845-851.

Miller, J.D. (**1989**). "Auditory-perceptual interpretation of the vowel," *J. Acoust. Soc. Am*. **85**, 2114-34.

Mullennix, J.W., Johnson, K., Topcu-Durgun, M. & Farnsworth, L.M. (**1995**). "The perceptual representation of voice gender," *J. Acoust. Soc. Am.* **98**, 3080-95.

Nearey, T.M. (**1978**). Phonetic Feature Systems for Vowels (IU Linguistics Club, Bloomington, IN).

Newman, R.S. (**1996**). "Individual differences and the perception-production link," *J. Acoust. Soc. Am.* **99**, 2592.

Nichols, R.C. & Bilbro, W.C. Jr. (**1966**). "The diagnosis of twin zygoticity," *Acta Genetica Med. Gemellol.* **16**, 265-75.

Nolan, F. & Oh, T. (**1996**). "Identical twins, different voices," *Forensic Linguistics* **3**, 39-49.

Nordström, P.-E. & Lindblom, B. (**1975**). " A normalization procedure for vowel formant data," in Proceedings of the 8th International Congress of Phonetic Sciences, Leeds, England.

Ostwald, P.F., Freedman, D.G., and Kurtz, J.H. (**1962**). "Vocalization of infant twins," *Folia Phoniatrica* **14**, 37-50.

Remez, R.E., Rubin, P.E., Pisoni, D.B. & Carrell, T.D. (1981) Speech perception without traditional speech cues. *Science* **212**, 947-950.

Shannon, R.V., Zeng, F.-G., Kamath, V. Wygonski, J. & Ekelid, M. (1995) Speech recognition with primarily temporal cues. *Science* **270**, 303-304.

Stafford, L. (**1987**). "Maternal input to twin and singleton children: Implications for langage acquisition," *Human Communication Research* **13**, 429-462.

Strand, E. A.(**1999**) Gender stereotypes and the perception of speech. OSU ms.

Syrdal, A. & Gopal, H. (**1986**). "A perceptual model of vowel recognition based on the auditory representation of American English vowels," *J. Acoust. Soc. Am.* **79**, 1086-1100.

Table 1. Sources of talker variation in speech.

Anatomical

- vocal folds - voice fundamental frequency, voice quality (breathy, creaky, etc.)
- vocal tract length - vowel formant frequency ranges
- vocal tract geometry - phonetic details, including patterns of gestural coordination

Behavioral

- native language - phoneme inventory, prosody, phonetic implementation
- dialect - phoneme inventory, prosody, phonetic implementation
- ideolect - ??

Table 2. The six pairs of twins who recorded for this study. Speakers QB & CK are dizygotic twins, the others are monozygotic.

speaker	**age**	**home (young)**	**home (current)**
BB	67	Longbottom, OH	Columbus, OH
BP			Columbus, OH
NJ	20	Columbus, OH	Columbus, OH
AJ			Columbus, OH
LW	26	Bronx, NY	New York, NY
PW			Columbus, OH
RJ	43	Ahoski, NC	Columbus, OH
AM			Columbus, OH
CN	21	Noble, OK	Ada, OK
MN			Ada, OK
QB	36	Norman, OK	Norman, OK
CK			Norman, OK

Table 3. The list of words for the recordings. The underlined words were used as stimuli in the perception experiments.

Steady-State Vowels	Diphthongs	Rhotic Vowels	VOT
heed	how'd	fear	pit
hid	hide	far	bit
head	void	fur	pie
had	aid	four	by
odd	owed	fair	peed
awed			bead
hud			putt
hood			but
who'd			

Figure Captions

Figure 1. Bias (β) and sensitivity (d') as a function of experiment.

Figure 2. Results of the multidimensional scaling analysis of the proportion of 'same talker' responses in experiment 3. Each point represents the location of a speaker in the listeners' perceptual space. Points representing twins are connected to each other by a line.

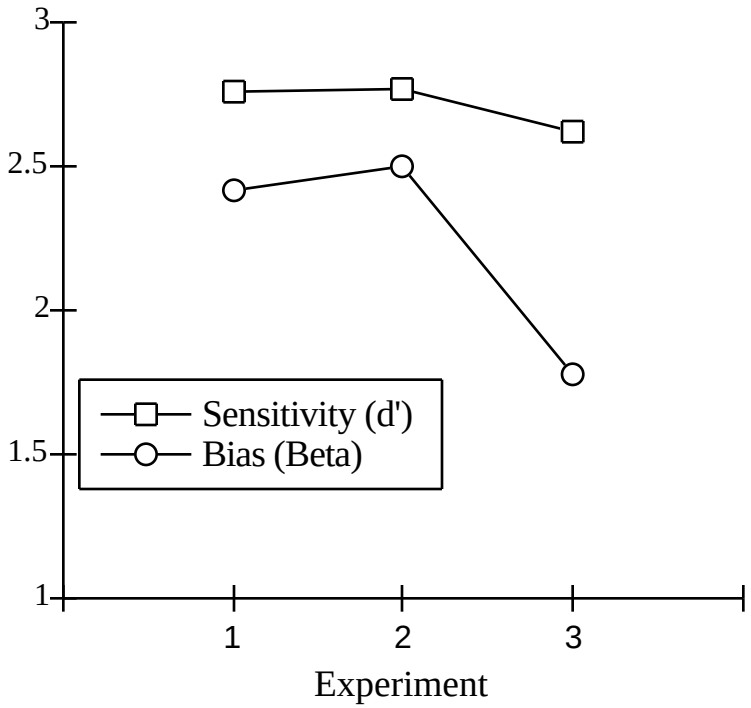

Figure 1.

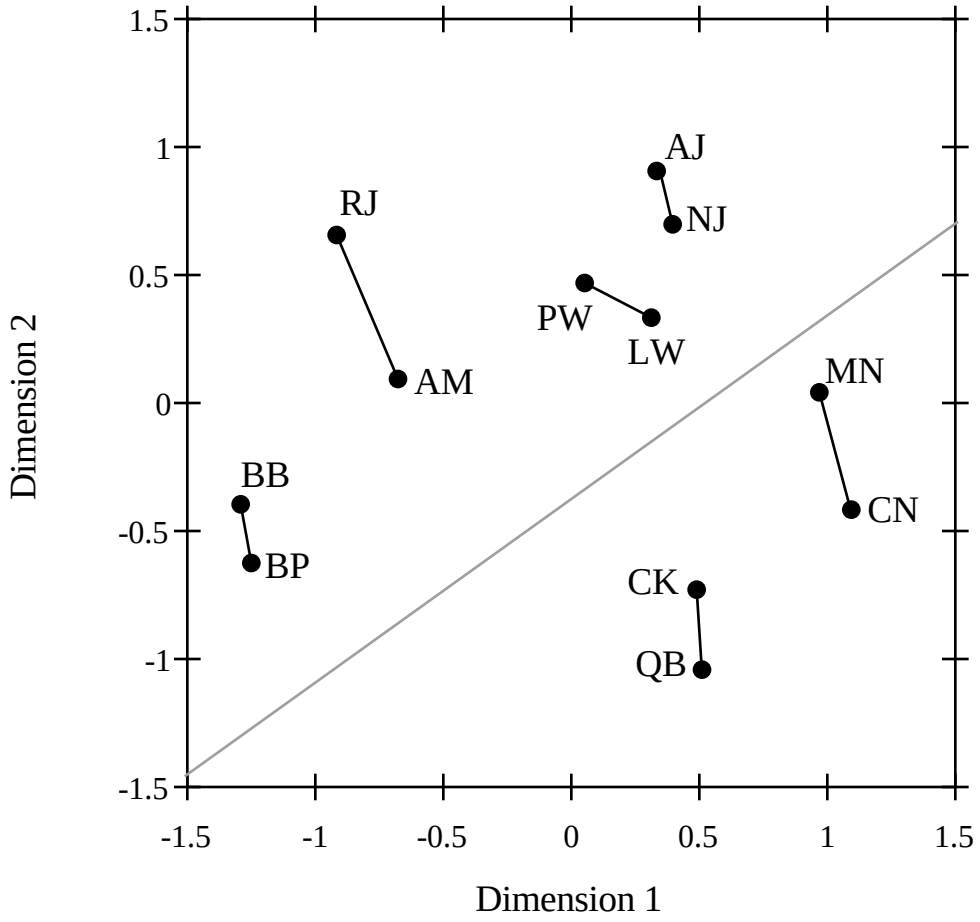
1.5
1
0.5
0
-0.5
-1
-1.5
Dimension 2
-1.5
-1
-0.5
0
0.5
1
1.5
Dimension 1
AJ
NJ
RJ
AM
PW
LW
MN
CN
BB
BP
CK
QB

Figure 2.