
10 Spoken Expression of Individual Identity and the Listener

Robert E. Remez

To a listener, a brief phrase is sufficient to evoke a personal impression of a talker. The speech of someone familiar can have a subtle impact, and even an unfamiliar voice suggests traits, true or false, of the one who spoke. It should not be surprising that qualitative variation in speech has commonly been identified as the source of these personal impressions, and a substantial technical literature reports investigations of the acoustic correlates of regional and social group, a talker's age and sex, and a talker's affective state and arousal (classics include Abercrombie, 1967; Bricker & Pruzansky, 1976; Hecker, 1971; a recent review is offered by Kreiman, Van Lancker-Sidtis, & Gerratt, 2005). However, the custom of ascribing impressions of personal quality solely to the extralinguistic properties of speech has led to neglect of a significant dimension of individual variation that can be critical to the expression and perception of personal identity. An individual's characteristic way of producing the consonants and vowels that compose words can be distinctive even when a spoken expression is otherwise unremarkable, prosaic, and normative. Although an individual's articulation of speech is a linguistically regulated aspect of communication, personal consistency in this aspect of symbolic expression provides a rich assortment of indexical attributes. Some recent and new studies draw attention to the potential role of *idiolect*, the individual linguistic characteristics nested within dialect, in individual identification.

To expose the argument and evidence, it is useful to consider the precise circumstances in which we speak and listen. A gloss of the production of language spotlights this condition (Levelt, 1989). In a conversation, an urge to take a turn initiates a semantic aim, an intention to create an utterance of a certain dimension. This goal initiates a compositional function to choose words and to assemble them into an order with representational properties that more or less match the semantic aim. There are many ways to choose words: by shade of meaning, perhaps, or also commonness or rarity, or by alliteration or rhyme. The dimensions of variation among the words of a language include semantic properties and differential incidence, and variation is well studied in each of these. Alliteration and rhyme are characteristics of the meaningless phonological form by which words are distinguished. Phonemic form amounts to an addressing scheme for items in

lexical memory, marking distinctions among words, although this dimension of variation also supplies a set of abstract criteria for articulating the word and making a private linguistic composition public.*

Through this last juncture in production, talkers and listeners within a community share the language forms that are used to compose utterances. This cheerful confluence of symbolic processes is compromised, however, by vocal dispersion when language is spoken. Classically, the population of a community is said to vary greatly in the vocal anatomy used to express the phonemic form of language phonetically (Fant, 1966; Ladefoged, 1967). Intuitively, this is evident in contrasting the voices of men, women, and children who, speaking the same words, can sound distinct. The speech of each individual gives sound to the shared linguistic forms in the unique characteristics of the person who spoke. In consequence, each individual expresses personal characteristics in an utterance.

The challenge to understand the perceptual analysis of individual expression nested within spoken language can be gauged in a recent project of talker identification by listening (Krauss, Freyberg, & Morsella, 2002). Thirty-nine opportunistically chosen strangers were recorded speaking two neutral test sentences. Each talker was also photographed, full body, against a neutral background, and was asked to report age, height, and weight. Listeners who were familiar with none of the talkers in the set were subsequently asked to identify each in a procedure that included listening to a speech sample and choosing between two photographs. One photo depicted the talker, the other a foil. Imposing the condition in which talkers were unfamiliar to listeners prevented listeners from succeeding in an identification task by remembering the vocal traits of specific individuals. Nonetheless, listeners identified talkers better than they would have by guessing alone. Which attributes of the voice did they detect? A second test calibrated a listener's ability to assess rough somatic proportions—the age, height, and weight—of talkers from the speech samples and from the photographs. Although judgments were more accurate from photographs than from speech samples, judgments from speech were still surprisingly well correlated with the actual age, height, and weight of the talkers who spoke them. Sampling the mix of attributes available in an individual's speech, listeners discern personal characteristics, and this chapter reviews the pertinent research terrain from a high altitude, observing the sources of distinctiveness in a talker's speech and the ways that a listener resolves these perceptually.

* This caricature of the production of an utterance might seem psychologically false, a form of assembly more apt for building a Ford than an expression. The psychological description will seem truer if the stages described sequentially here are understood as nested formal requirements satisfied in a hybrid function, part parallel. Whether the constituents are assembled in advance completely or on the fly, linguistic form is shared throughout a community of talkers, while the physical means of expression is unique to each talker and situation.

PERSONAL AND VOCAL IDENTITY

Within the technical literature on the perception of linguistic aspects of speech, individual variation among talkers has commonly been treated as a nuisance (Halle, 1985). Because the linguistic components of any message must be roughly the same for each communicating talker and listener, from this perspective the acoustic effects of the unique embodiment of each individual count as noise in a transmission line. Accordingly, linguistic perception has been attributed to the detection of canonical acoustic correlates of phonemes that are present in the speech stream regardless of the talker or the conditions under which speech was produced. In complementary fashion, the projects to describe the perceptible acoustic signature of individual talkers have largely sought to distill the durable aspects of an individual's speech that transcend the linguistic properties expressed moment to moment. In this conceptualization, the speech of an individual is treated as a composite of linguistic and personal properties, each with different acoustic correlates and each obliging a different perceptual function. Whether the goal has been identification of individuals by ear, by algorithm, or by visual inspection of spectrograms, researchers have tended to define the problem as a kind of commerce with nonlinguistic acoustic attributes as currency (Bricker & Pruzansky, 1976). In the past few years, though, the boundary between linguistic and personal aspects of speech production has been redrawn empirically. Here, a rendition of linguistic contributions to talker perception is offered; for a review of qualitative influences in the perception of spoken words, see Nygaard (2005).

To begin at the beginning, a talker can be distinctive under conditions in which an aspect of speech or constellation of aspects is unique, is produced consistently, and is resolvable perceptually by listeners. In this regard, listeners are apparently voracious for useful properties of speech, and while there is apparently no single attribute that counts as a universal indexical tag, there is a large variety that is possibly valuable. Truly, the catalog is large, and a synopsis of this literature can readily be found elsewhere (Bricker & Pruzansky, 1976; Hecker, 1971; Kreiman, 1997; Remez, Fellowes, & Rubin, 1997). Although some of the targeted properties of speech are direct consequences of vocal anatomy and physiology, others are an outcome of linguistic exposure and social role.

ANATOMICAL VARIATION AMONG TALKERS

Some aspects of individual variation seem ineluctable. The average and range of vocal pitch is associated with the mass of the larynx (Ishizaka & Flanagan, 1972). The frequency range and central tendency of the natural resonances of the vocal tract are determined by the scale of the supralaryngeal vocal tract (Fant, 1960). These characterizations are among the oldest in acoustic phonetics, appearing in the foundational monograph by Joos (1948), and the observations have had wide conceptual influence. Despite large overlap throughout the range, there are consistent

differences in average fundamental frequency and glottal spectrum over the life cycle and between male and female talkers. However, variation in fundamental frequency falls short as a marker of individual identity, at least considering average frequency, frequency excursion, and most comfortable frequency, indicating that the functional states of the larynx vary more than the anatomy. For instance, there are differences in characteristic fundamental frequency reported throughout the daily cycle (Garrett & Healey, 1987) and given the brief period of these cycles, they surely must be accountable without appeal to variation in laryngeal size.

Similarly, vocal resonances vary with the length and shape of the vocal tract. Sound production governed by the linguistic properties of an utterance creates modulation of vocal resonances that constitutes the acoustic stream of speech for a listener, and the unique scale of a talker determines the frequency composition of the stream. Although vocal tract length is only weakly correlated with bodily stature, and although there is significant overlap throughout the range, again, to a first approximation, adult males tend to exhibit the lowest resonant frequencies, adult females a bit higher, and children the highest (Goldstein, 1980; Peterson & Barney, 1952). Under some conditions, perceivers act as if they were normalizing the formant variation attributable to differences in vocal tract scale (Ladefoged & Broadbent, 1957). In other words, perceptually calibrating the scale of a talker influences the apparent linguistic properties of utterances. This evidence of perceptual sensitivity to the precise dimensions of variation once suggested that vocal tract scale is a primary indexical attribute of talkers (see Pisoni, 1997). As in the case of fundamental frequency, though, the average vocal spectrum offers less in fact than in principle as an indicator of identity. One reason is that the resonant frequency is an indirect measure of vocal tract length, varying with the rounding and spreading of the lips, and the height of a mobile jaw and larynx (Fant, 1960). In similar circumstances to the larynx, the possibility of many functional states available within a single vocal tract opposes a simple conversion of sound to anatomical scale.

PHYSIOLOGICAL VARIATION AMONG TALKERS

Many degrees of freedom are available for creating qualitative acoustic variation laryngeally and supralaryngeally. Neither the exact dimensionality nor the physiology of production is understood thoroughly, although the literature contains a few meta-analyses that aim to define parameters of this multidimensional aspect of the voice. In one landmark review, Gelfer (1988) justified seventeen dimensions of qualitative variation in the voice, admittedly based in part on clinical evaluation of dysphonia and other pathology. Table 10.1 reproduces the dimensions derived in this study, and by using each of its binary dimensions as a Likert scale it is possible to parameterize the qualities of a speech sample, or of a specific talker's voice. None of the dimensions creates linguistic contrasts, at least not in English, a language often examined by researchers. The dimensions of variation can be considered aspects of personal style.

The use of the larynx to produce differences in vocal quality has a counterpart in the vocal tract. In one analytic approach to this, individual talkers are

TABLE 10.1
Dimensions Derived by the Meta-analysis of
Gelfer (1988) for Describing the Qualitative
Characteristics of the Speaking Voice

High pitch – Low pitch	Clear – Hoarse
Loud – Soft	Unforced – Strained
Strong – Weak	Soothing – Harsh
Smooth – Rough	Melodious – Raspy
Pleasant – Unpleasant	Breathy – Full
Resonant – Shrill	Nasal – Not nasal
Animated – Monotonous	Young – Old
Steady – Shaky	Slow – Rapid
Liked – Disliked	

characterized as adopting potentially distinctive vocal postures, each establishing a neutral articulatory state. The articulation required to produce linguistically governed gestures is constrained by assuming a consistent vocal posture of a lowered larynx, for instance, or a pharyngealized station for the tongue body, and such postural adjustments are understood to modulate the shape of the long-term integrated spectrum of speech (Nolan, 1983). A listener who is able to abstract the typical vocal spectrum of familiar talkers might then be able to recognize one from a speech sample of sufficient duration. The psychological question raised by this theory of qualitative variation is whether recognizing a supralaryngeally caused timbre* permits identification of an individual, or whether individual identity and individual style are resolved together. This aspect of individual identification deserves closer attention, although the evidence in hand does not encourage

* Timbre, or sound quality, is a well-established descriptive topic within auditory sensory psychology, and a classic review by Hirsh (1988) makes it possible to gauge the difficulty of creating an account of individual identification by resolution of timbre. The term itself refers to the qualitative dimension of auditory experience, and it is defined by exclusion. Namely, when two sounds of identical pitch and loudness—here, the terms refer to psychological states, not to the physical properties of frequency and power—can be differentiated by their individual qualities, the dimension of difference is timbre. Of course, timbre is a constant facet of auditory experience, and there are many studies of the acoustic causes of different timbre impressions. Hirsh notes, though, that descriptions of timbre are readily asserted, including some in Table 10.1, that pertain to the mechanical means of production rather than to the sound quality that is experienced. In other words, some claims conflate attributes of sensory quality with attributes of the object that produces it, describing a sound as “mechanical,” or “oboe-like,” for instance. It would be tautology to assert that the musical instruments oboe and saxophone are identified by an experience of “oboe-like” and “saxophone-like” timbre, or that a particular individual is identified by recognizing his pleasant, breathy timbre. Although attributes are subordinate to the individual who expresses them, auditory quality is sensory, primary, and uninterpreted, while vocal quality is an aspect of object perception.

the conceptualization of a listener first sampling an unknown voice to resolve the characteristic long-term vocal quality of the talker, and then identifying the person who spoke. The interval required to estimate an individual's pitch range or average vocal quality has temporal coordinates inconsistent with the fluency of perception (Pollack, Pickett, & Sumbly, 1954). A second of speech might be all that is required to recognize a familiar talker, hardly a sufficient span to compile a sensory assay of variation in voice pitch and quality.

These cautions in mind, there is much evidence of the usefulness of attention to something correlated with glottal period and spectrum or consistent supralaryngeal habits, in the perceptual identification of individual talkers. The brevity of samples that evoke personal impressions, as little as 1 s, argues against a perceptual norming operation in which large-scale distributional characteristics of glottal and supralaryngeal effects are assessed preliminary to identification and then are compared to remembered characteristics of talkers who are familiar to the listener. Yet, the key features associated with the perception of voice quality and the acuity of perceptual resolution are yet to be established empirically. The facility of listeners identifying a talker from whispered samples is impaired relative to phonated speech (Pollack et al., 1954; Tartter, 1991; cf., Eklund & Traunmüller, 1997), evidence of a role of glottal period and spectrum in talker identification by ear. Supralaryngeal characteristics play a role, as shown in the successes, albeit at reduced accuracy, in the identification of talkers using an electrolarynx to substitute for the natural voice source (Coleman, 1973).

More challenging is the evidence that a talker who produced speech that was temporally reversed and presented for naming remains identifiable in many instances (Clarke, Becker, & Nixon, 1966; Van Lancker, Kreiman, & Emmorey, 1985), proof that lexical and some phonetic attributes are not necessary for the perceptual identification of talkers. In such cases, the glottal and supralaryngeal sources of quality are preserved in syllable nuclei and in consonant spectra of long duration, including fricatives, nasals, and the hold portion of liquids. The time-critical evolution of stop consonants and affricates is grossly disrupted, indicating that some critical features pertinent to a talker's traits can be available without these linguistic attributes.

LINGUISTIC VARIATION AMONG TALKERS

The articulators are a convergence point in expression. Linguistic properties might govern the abstract criteria for articulating the word *heet*, *hoot*, or *hot*, but a talker's age, sex, physique, vitality, attitude, and distance from the listener determine qualitative aspects of production. At least, the common assumption has been that speech incorporates all of these influences as such. Nonetheless, it would be false to conclude that the average spectral difference between talkers mostly reflects the differences due to the length of the supralaryngeal vocal tract and the language-independent bodily characteristics of specific individuals. A couple of studies indicate the nature of this complex relation.

The evidence that addresses this directly includes a cross-language comparison of the average difference in male and female formant frequency in the long vowels of twenty-six languages (Johnson, 2005). In some languages, male and female average spectra differed greatly, as if variation in resonant frequencies across talkers was the simple consequence of differences in the scale of the vocal anatomy. In this class were Wari', an Amazonian language, California English, spoken in many parts of California, and Russian. Yet, in other languages, the formant frequencies typical of one sex were highly similar to those typical of the other. In this class were Danish, spoken in northern Europe, Angami, spoken in South Asia, and Paicî, spoken in New Caledonia. This sort of comparison offers a fresh look at the interaction of linguistic and personal characteristics.

Despite differences between individuals in the anatomy of articulatory structures, the expression of speech allows wide range, and not just extralinguistically. In a hypothetical push and pull between linguistic and extralinguistic regulation of speech, this analysis of vowel resonant frequencies shows that the boundary between the two domains is movable, as if attributes of speech that are linguistically conditioned in one environment are free to vary in another. Evidently, a language can regulate phonetic compensation for intrinsic anatomical differences. This might affect vocal quality and hypothetically imperil individual identification based on this acoustic property for speakers of Danish but not Russian, although the pertinent data have not been produced.

There is evidence of other ways to mark the sex of a talker phonetically than with the height, advancement, and rounding of vowels. In one analysis, the disposition to hold or to release the final consonant of a closed syllable at the end of a sentence—for instance the *t* in *just*, the *d* in *pursued*, the *k* in *work*—was correlated with the sex of the talker (Byrd, 1994). Other evidence of linguistic and personal interchange is also observed in American English. Specifically, sex differences in vowel formant frequency are reported in juvenile subjects, well before the age of expression of secondary dimorphism in the dimensions of larynx and vocal tract (Lee, Potamianos, & Narayanan, 1999; Perry, Ohde, & Ashmead, 2001). Of course, if a talker is obliged linguistically to regulate a specific dimension of production, then when the talker shifts into the role of perceiver this aspect of speech is potentially resolvable. In addition, if it is perceptually resolvable for self-regulation, the same capacity is probably available, at least potentially, as a propensity to notice such variation in the speech of others.

Apart from apprehending the message, though, an ordinary listener—which designation includes this author as well as the reader of this chapter when we leave the laboratory—might be aware only vaguely of sex-linked phonetic expression, experiencing it as a phenomenon of character and not of phonology. Indeed, if it is a consistent feature of a talker's expression, it is no less an aspect of self than of lexical addressing. Although talker and listener alike might attend to this aspect of articulation without explicit awareness of the linguistic patterning involved, at subordinate levels of control the phonetic forms engaged in personal marking must be registered in precise detail. Without exact registration of the forms undergoing the alternation that marks sex while pointing to

words, there is no consistency in production. However the compromise is struck between the shared symbolic forms and the phonetic assortment available for a specific individual's expression, a talker must produce and a listener must apprehend a legitimate allophone of a phoneme contrast in order to distinguish an intended word from other words of like composition and class (Luce, 1986; Luce & Pisoni, 1998).

The intrusion of language on the pure expression of natural bodily differences is also a feature of dialect and arguably of idiolect. In dialect, local communities drift together; in idiolect, an individual drifts distinctively, producing only a subset of the characteristics definitive of a dialect. Whether anatomy guides an idiosyncrasy of production, due to the shape and mobility of articulatory structures, or whether it is freely assigned, is a topic deserving attention of researchers. However, a provocative and informative instance is reported by Johnson and Azara (2000). In a study of the speech of identical twins, they reported that many of the utterances in the samples were so similar, comparing one twin to the other, that the interaction of linguistic exposure and biological potential crystallized largely in the same way in each. The differences they reported were no greater than is typical of a single individual producing different versions of the same utterances. Yet, twins also differed from each other consistently in other formant measures (see also Nolan & Oh, 1996), as if a subset of contrasts specifically marked the difference between the twins within a larger set of shared elements. This study offers a glimpse of the gradient from assimilation of the community standard and individual defection in rough, if not strict, analogy to social acts of different scale (see Pickett & Brewer, 2001). When an individual takes on characteristics of the dialect, sharing features of production with the local community, this minimizes individuality in phonetic expression. When an individual differs from the group extensively, this distinctiveness is purchased at the cost of easy phonetically based identification of community membership.

The research required to refine these notions with evidence would examine the interplay between dialect, idiolect, and the lexicon (for example, Clopper & Pierrehumbert, 2008). For this purpose, it would be ideal to conduct a longitudinal study of language development of monozygotic dectuplets reared together, in order to contrast inheritance, linguistic exposure to a community, and the resolution of conflicting dispositions to match the community dialect and to establish a unique idiolect within it. Failing to meet this methodological objective, there are other more approximate means of examining how language highlights and perhaps exaggerates differences between individuals in the composition of an individual's phonetic inventory.

PHONETIC IDENTIFICATION OF TALKERS

Some studies have attempted to provide a critical empirical test of the central claim that consistency in phonetic production functions indexically, independent of extralinguistic somatically correlated acoustic properties of speech (Remez et al., 1997). The method relied on tests of perceptual identification in which

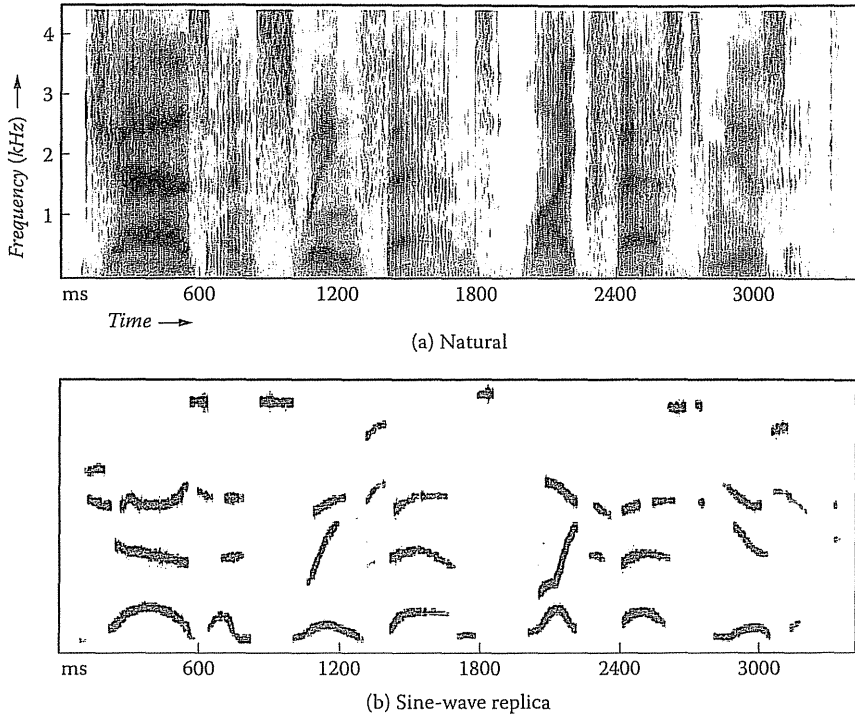


FIGURE 10.1 A comparison of the spectral characteristics of natural speech and a sine-wave replica. (a) A spectrographic analysis of the natural sentence: Jazz and swing fans like fast music. Note the broadband resonances, aperiodic constituents, and glottal pulses. (b) A spectrographic analysis of a sine-wave replica of the natural utterance. The coarse-grain spectrotemporal pattern of the natural model is preserved despite the use of time-varying sinusoids and the absence of the natural acoustic products of vocalization. From Pardo & Remez (2006).

speech was presented in synthesis. Rather than using traditional speech synthesis, which can incorporate acoustic properties that evoke natural vocal quality, the tests used drastically reduced spectral patterns consisting solely of three or four time-varying sinusoids, each set to replicate the frequency and amplitude properties of a resonance in a natural utterance. Such patterns, often designated as sine-wave speech, lack an acoustic component that exhibits the variation in fundamental frequency; it is simply omitted from the synthetic acoustic pattern. Likewise, nothing of the glottal spectrum remains, nor is the overall impoverished spectral shape more than an insinuation of the original natural utterance on which the synthetic version is modeled. Neither are the tone components related harmonically, nor are the fricative portions aperiodic, nor are the resonances broadband (See Figure 10.1). Despite intelligibility evoked by coherent

spectrotemporal variation, natural vocal quality is utterly lost. These odd sounds, so different from the natural utterances on which they were based, should not have evoked an impression of an identifiable talker if the acoustic attributes that matter indexically are solely qualitative in origin and effect.

One test used ten different talkers and two groups of listeners. One listener group was not familiar with any of the original talkers, and they were asked to determine which of two sine-wave sentences had been produced by the talker who spoke a natural sentence also presented on each test trial. A second group had become familiar over a span of decades with each of the talkers whose speech provided natural models for the sine-wave versions. These listeners were asked to identify the talkers by name from remembered characteristics. Although there was a slight performance benefit exhibited by those who knew the talkers from ordinary interchange, both groups of listeners were able to identify the talkers producing the sine-wave samples without relying on vocal pitch and quality. Moreover, frequency transposition of the sine-wave constituents that eliminated acoustic correlates of vocal tract scale variation did not prevent identification of sine-wave talkers (Fellowes, Remez, & Rubin, 1997). Most telling were the errors of identification. Listeners seemed to disregard overall speech rate, a property of natural speech preserved in sine-wave sentences, and they tended to ignore the acoustic correlates and perceptual consequences of sex differences among the sine-wave talkers. Instead, it seemed as if sine-wave talkers were confused for one another when they exhibited similar segmental phonetic details, which is to say, when talkers used similar phonetic variants of consonants and vowels.

In another related study, a procedure of direct estimation of similarity showed that the pattern of perceptual contrasts among talkers was much the same whether talkers were presented as natural samples or as synthetic sine-wave versions lacking natural vocal quality (Remez, Fellowes, & Nagel, 2007). By asking listeners to weigh the attributes of talkers rather than the sound of the samples, the undeniable qualitative differences between natural and sine-wave versions were reduced in salience, presumably in favor of the subphonemic phonetic variants that remain when natural qualitative attributes are diminished or expunged.

Although the cases that produced this empirical proof used electroacoustic methods to eliminate natural vocal quality, there are more prosaic causes of qualitative distortion that probably spare the perceptual effects of linguistically controlled speech production. Phonetic patterns are simple to maintain when typical qualitative production is disrupted in laryngitis or when naturally produced qualitative effects are lost due to a poor transmission line, perhaps in an intercom or telephone. Because phonetic properties survive qualitative distortion robustly, the ability of a listener to identify a talker would gain durability if perception tracked phonetic inventory in addition to vocal quality. Certainly, this conclusion is encouraged by the mundane experience of recognizing a talker despite distortion imposed by a typical telephone, which filters the speech by band limiting the frequency range, and scrambles the phase relations among the spectral components.

A SIDEBAR ON IDENTITY

Identity is an intriguing and nuanced topic. Yet, a narrow forensic focus has governed a large portion of the research on talker identification, even when a scientific project did not entail identifying an unknown talker, or verifying that a specific detainee had self-identified honestly. The consistent axiom of this approach is the permanence of identity from cradle to grave and beyond. The forensic presumption that an individual's identity is fixed is somewhat justified by the limited time span of a legal proceeding, although the volatility of character has been a significant topic within social science since the discovery of the susceptibility of traits to circumstantial modulation. Or, perhaps longer:

Even sound authors are wrong in stubbornly trying to weave us into one invariable and solid fabric.... Anyone who turns his prime attention onto himself will hardly ever find himself in the same state twice. I give my soul this face or that, depending on which side I lay it down on. I speak about myself in diverse ways: that is because I look at myself in diverse ways. Every sort of contradiction can be found in me, depending upon some twist or attribute: timid, insolent, chaste, lecherous; talkative, taciturn; tough, sickly; clever, dull, brooding, affable; lying, truthful; learned, ignorant; generous, miserly and then prodigal—I can see something of all that in myself, depending on how I gyrate; and anyone who studies himself attentively finds in himself and in his very judgment this whirring about and this discordancy. There is nothing I can say about myself as a whole simply and completely, without intermingling and admixture. (de Montaigne, 1574; translated and edited by Screech, 1991, pp. 373–377.)

The changes in speech that accompany these varied aspects of the self are rarely examined with the precision required to say that presenting one or another facet of character is expressed both qualitatively and linguistically. From partial treatments, it is possible to see that if this conceptualization is fair, then research that adopts the premise of immutable identity conveyed qualitatively will not make useful measures. At least part of an individual's phonetic repertoire is tied to an alternation in diction associated with social roles and registers (Labov, 1986). It is also plausible that mood is marked allophonically, although research has chiefly concerned the consequences of affect on resonance spectrum, a result of the articulation by the lips with cheer and gloom (Tartter, 1980). One alternative notion is that the expression of mood is actually performed in part, and is not simply a somatically necessary perturbation of vocal expression. If we accept the conclusion of research on clear and casual speech (Picheny, Durlach, & Braida, 1985), namely, that it is signaled in part by such alternations as /drd##ju/ → /drdʒə/, then a permissible extrapolation is that other alternations in attitude might also be conveyed in the subphonemic pattern of expression in addition to qualitative effects.

One necessary consequence of a link between self-expression and phonetic expression is a perceptual cost that is already gauged, and this is the typical effect on individual perception of a known talker's use of a vocal disguise (Hollien, Majewsky, & Doherty, 1982). Even listeners who are capable of identifying a specific individual can be readily fooled when that talker attempts to mislead by

affecting an uncharacteristic manner of speech. What is the source of the cost, perceptually, that results in erroneous identification? In affecting a vocal disguise, a talker might present a rare self, counterfeit or merely unfamiliar to the listener, and such changes conceivably precipitate a different allophone repertoire as well as unfamiliar qualitative mannerisms, even though disguised and undisguised talker are the same person.

CONCLUSIONS

A few simple ideas about speech production frame this chapter, which has summarized their defense, empirically and conceptually, and noted the perceptual consequences. It is fair to state them plainly as a means to identify the improvements that research can bring to this approach:

- An individual expresses personal identity in speech.
- Anatomy and physiology determine some acoustic characteristics of speech.
- Linguistic experience regulates expression in dialect and idiolect.
- Mood, motive, and situation affect the variety of linguistic phonetic properties.
- A listener often falsely ascribes dialectal and idiolectal marking to qualitative attributes of an individual talker.
- Linguistic differences among individuals persist across qualitative variation.

These premises are yet to be fully secured by evidence in the argument that talkers are consistent in their idiosyncratic allophonic habits and that these linguistic properties evoke impressions of personal character in listeners. Yet, there is a last argument to review here that is pertinent to the plausibility of this conceptualization. The self-regulatory goal of an individual talker poses a conflict for each language learner: to talk enough like the group to be taken as a member, yet to reserve some of the expressive potential of the linguistic phonetics to be unique within the group. Implicitly, this problem requires a talker to use phonetic perception to calibrate and to regulate articulation, fixing the center and range of the linguistic community and the centrality or eccentricity of the self—or selves. Once this capacity is developed, the perceptual sensitivity required to meet this challenge of adaptive self-regulation, linguistically, is thereafter available for another use: to calibrate the attributes of other talkers by their phonetic characteristics as well as their vocal quality.

Now, if we could only explain why it is often a pleasure to notice these subtle details about each other...

ACKNOWLEDGMENT

For many provocative, enlightening, tough, and patient discussions over a span of decades, the author is grateful to Bob Krauss. For support of this research, thanks are extended to the National Institute on Deafness and Other Communication Disorders (award DC00308).

REFERENCES

- Abercrombie, D. (1967). *Elements of general phonetics*. Chicago: Aldine.
- Bricker, P. D., & Pruzansky, S. (1976). Speaker recognition. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics* (pp. 295–326). New York: Academic Press.
- Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech Communication*, 15, 39–54.
- Clarke, F. R., Becker, R. W., & Nixon, J. C. (1966). Characteristics that determine speaker recognition. *Electronic Systems Division, Air Force Systems Command Report ESD-TR-66-638*. Hanscom Field, MA.
- Clopper, C. G., & Pierrehumbert, J. L. (2008). Semantic context and the Northern cities chain shift. *Journal of the Acoustical Society of America*, 124, 1682–1688.
- Coleman, R. O. (1973). Speaker identification in the absence of inter-subject differences in glottal source characteristics. *Journal of the Acoustical Society of America*, 53, 1741–1743.
- de Montaigne, M. (1574). On the inconstancy of our actions. [translated by M. A. Screech, 1991, *The essays of Michel de Montaigne, II, 1*: (pp. 373–377). London: Allen Lane.
- Eklund, I., & Traunmüller, H. (1997). Comparative study of male and female whispered and phonated versions of the long vowels of Swedish. *Phonetica*, 54, 1–21.
- Fant, C. G. M. (1960). *The acoustic theory of speech production*. The Hague: Mouton.
- Fant, C. G. M. (1966). A note on vocal tract size factors and nonuniform *F*-pattern scalings. *Speech Transmission Laboratory Quarterly Progress and Status Report*, 4, 22–30. Stockholm, Sweden: Royal Institute of Technology.
- Fellowes, J. M., Remez, R. E., & Rubin, P. E. (1997). Perceiving the sex and identity of a talker without natural vocal timbre. *Perception & Psychophysics*, 59, 839–849.
- Garrett, K. L., & Healey, E. C. (1987). An acoustic analysis of fluctuations in the voices of normal adult speakers across three times of day. *Journal of the Acoustical Society of America*, 82, 58–62.
- Gelfer, M. P. (1988). Perceptual attributes of voice: Development and use of rating scales. *Journal of Voice*, 2, 320–326.
- Goldstein, U. (1980). An articulatory model for the vocal tracts of growing children. Unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.
- Halle, M. (1985). Speculations about the representation of words in memory. In V. A. Fromkin (Ed.), *Phonetic linguistics* (pp. 101–114). New York: Academic Press.
- Hecker, M. H. L. (1971). Speaker recognition: An interpretive survey of the literature. *ASHA Monographs*, No. 16, 1–103.
- Hirsh, I. J. (1988). Auditory perception and speech: In R. C. Atkinson, R. J. Herrnstein, G. Lindzey, & R. D. Luce (Eds.), *Stevens' handbook of experimental psychology, Volume I: Perception and motivation* (pp. 377–408). New York: Wiley-Interscience.
- Hollien, H., Majewsky, W., & Doherty, E. T. (1982). Perceptual identification of voices under normal, stress and disguise speaking conditions. *Journal of Phonetics*, 10, 139–148.
- Ishizaka K., & Flanagan J. L. (1972). Synthesis of voiced sounds from a two-mass model of the vocal cords. *Bell System Technical Journal*, 116, 1233–1268.
- Johnson, K. (2005). Speaker normalization in speech perception. In D. B. Pisoni & R. E. Remez (Eds.), *Handbook of speech perception* (pp. 363–389). Oxford, UK: Blackwell.
- Johnson, K., & Azara, M. (2000). The perception of personal identity in speech: Evidence from the perception of twins' speech. Unpublished manuscript.
- Joos, M. (1948). Acoustic phonetics. *Language*, 24(Supp.), 1–137.

- Krauss, R. M., Freyberg, R., & Morsella, E. (2002). Inferring speakers physical attributes from their voices. *Journal of Experimental Social Psychology*, 38, 618–625.
- Kreiman, J. (1997). Listening to voices: Theory and practice in voice perception research. In K. Johnson & J. W. Mullenix (Eds.), *Talker variability in speech processing* (pp. 85–108). San Diego, CA: Academic Press.
- Kreiman, J., Vanlancker-Sidtis, D., & Gerratt, B. R. (2005). Perception of voice quality. In D. B. Pisoni & R. E. Remez (Eds.), *Handbook of speech perception* (pp. 338–362). Oxford, UK: Blackwell.
- Labov, W. (1986). Sources of inherent variation in the speech process. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 402–425). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Ladefoged, P. (1967). *Three areas of experimental phonetics*. London: Oxford University Press.
- Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America*, 29, 98–104.
- Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *Journal of the Acoustical Society of America*, 105, 1455–1468.
- Levelt, W. J. M. (1989). *Speaking*. Cambridge, MA: The MIT Press.
- Luce, P. A. (1986). A computational analysis of uniqueness points in auditory word recognition. *Perception & Psychophysics*, 39, 155–158.
- Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, 19, 1–38.
- Nolan, F. (1983). *The phonetic bases of speaker recognition*. Cambridge: Cambridge University Press.
- Nolan, F., & Oh, T. (1996). Identical twins, different voices. *Forensic Linguistics*, 3, 39–49.
- Nygaard, L. C. (2005). Perceptual integration of linguistic and nonlinguistic properties of speech. In D. B. Pisoni & R. E. Remez (Eds.), *Handbook of speech perception* (pp. 390–413). Oxford, UK: Blackwell.
- Pardo, J. S., & Remez, R. E. (2006). The perception of speech. In M. Traxler & M. A. Gernsbacher (Eds.), *The handbook of psycholinguistics* (2nd ed., pp. 201–248). New York: Academic Press.
- Perry, T. L., Ohde, R. N., & Ashmead, D. H. (2001). The acoustic bases for gender identification from children's voices. *Journal of the Acoustical Society of America*, 109, 2988–2998.
- Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, 24, 175–184.
- Picheny, M. A., Durlach, N. I., & Braida, L. D. (1985). Speaking clearly for the hard of hearing. I: Intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research*, 28, 96–103.
- Pickett, C. L., & Brewer, M. B. (2001). Assimilation and differentiation needs as motivational determinants of perceived ingroup and outgroup homogeneity. *Journal of Experimental Social Psychology*, 37, 341–348.
- Pisoni, D. B. (1997). Some thoughts on "normalization" in speech perception. In K. Johnson & J. W. Mullenix (Eds.), *Talker variability in speech processing* (pp. 9–32). San Diego, CA: Academic Press.
- Pollack, I., Pickett, J. M., & Sumbly, W. H. (1954). On the identification of speakers by voice. *Journal of the Acoustical Society of America*, 26, 403–406.
- Remez, R. E., Fellowes, J. F., & Nagel, D. S. (2007). On the perception of similarity among talkers. *Journal of the Acoustical Society of America*, 122, 3688–3696.

- Remez, R. E., Fellowes, J. M., & Rubin, P. E. (1997). Talker identification based on phonetic information. *Journal of Experimental Psychology: Human Perception and Performance*, 23, 651-666.
- Tartter, V. C. (1980). Happy talk: Perceptual and acoustic effects of smiling on speech. *Perception & Psychophysics*, 27, 24-27.
- Tartter, V. C. (1991). Identifiability of vowels and speakers from whispered syllables. *Perception & Psychophysics*, 49, 365-372.
- Van Lancker, D., Kreiman, J., & Emmorey, K. (1985). Familiar voice recognition: Patterns and parameters. Part 1: Recognition of backward voices. *Journal of Phonetics*, 13, 19-38.