Speaker Perception and Social Behavior:
Bridging Social Psychology and Speech Science

Robert M. Krauss and Jennifer S. Pardo

Columbia University

Keywords

Language and identity
Social interaction
Speaker perception
Speech accommodation
Speech acoustics
Speech perception
Speech variation
Vocal communication

Authors

Robert M. Krauss is a social psychologist whose research has focused on human communication with a particular emphasis on language and speech. He received his Ph.D. from New York University in 1964. After a stint as a research psychologist at Bell Telephone Labs, he taught at Princeton, Harvard and Rutgers Universities before coming to Columbia in 1970. He is the author (with C.-y. Chiu) of "Language and Social Behavior" in the 4th edition of *Handbook of Social Psychology* and (with Susan Fussell) of "Social Psychological Models of Interpersonal Communication" in *Social Psychology: A Handbook of Basic Principles*.

Jennifer S. Pardo received her Ph.D. in cognitive psychology from Yale University in 2000. During her tenure there, she was affiliated with Haskins Laboratories and continues to collaborate on research in psycholinguistics. In 2002, she began a post-doctoral research fellowship in the Human Communication Laboratory at Columbia University. She is a co-author (with R. Remez et al., 1994) of "On the perceptual organization of speech" in *Psychological Review*.

Abstract

Language plays a critical role in social life, and has become an important area of social psychological research. However, social psychologists have focused on the semantic-pragmatic levels of linguistic analysis, and have paid considerably less attention to the organized sound system that underlies speech. We distinguish between speech perception, which includes the processes underlying comprehension of the linguistic content of speech, and speaker perception, which includes effects of variability in speech that is not linguistically significant. The latter deals with phenomena that lie at the heart of social psychology. We describe two broad research areas that illustrate the insights a consideration of the phonological level of speech can contribute to an understanding of social life.

Speaker Perception and Social Behavior:
Bridging Social Psychology and Speech Science

Robert M. Krauss and Jennifer S. Pardo

Columbia University

In recent years, social psychologists increasingly have come to appreciate the role that language plays in social life. For the discipline, the consequences of this developing awareness have been salutary. Language is critically implicated in many of the core phenomena social psychologists study: causal attribution (Semin & Fiedler, 1991), social identity (Giles, Taylor & Bourhis, 1977; Lambert, Hodgson, Gardner & Fillenbaum, 1960), status and intimacy (Brown & Gilman, 1960; Holtgraves & Yang, 1990), and interpersonal relations (Giles, Mulac, Bradac & Johnson, 1987), to cite but a few. Taking the role of language into account has greatly enhanced our understanding of these important phenomena. In addition, because stimulus and response in social psychology are often verbal in form, many fundamental questions of methodology turn on issues that are implicitly linguistic (Bless, Strack, & Schwarz, 1993; Schwarz, Strack, Hilton, & Naderer, 1991).

With a few notable exceptions, when social psychologists have considered language they have focused on the semantic-pragmatic levels of linguistic analysis. Much less attention has been paid to the system of sound production that allows semantic representations to be transformed into the perceptually accessible form we call *speech*.[1] Language can be defined as an abstract set of principles that specify the relation of a sequence of sounds to a sequence of meanings. Social psychology has been concerned mainly with the sequence of meanings. To oversimplify somewhat, social psychologists have limited themselves to aspects of linguistic representations that are preserved in text; they have largely ignored the additional information that speech conveys. This is unfortunate for many reasons, not the least of which is that speech, in addition to its semantic content, contains information that bears directly on phenomena that lie at the heart of social psychology.

It is useful to distinguish between two related areas of investigation that involve speech processing: research on *speech* perception and *speaker* perception. Speech

perception research studies the process by which listeners extract linguistically significant information from acoustic input. The process is complicated by the fact that spoken language is both highly variable in its production and remarkably stable in its perception. The central issue is to understand how listeners derive a stable percept from such variable input. In contrast, speaker perception research studies the effects of the variability in speech that is not linguistically significant (in the sense in which we are using that term), but is neither arbitrary nor idiosyncratic. An example may help to clarify the distinction. In American English, the height of the vowel in a word like "caught" may vary considerably from speaker to speaker, including pronunciations that range at the extremes from /kAt/ (a homophone of "cot"), to /k„t/ (pronounced "cawt"), to /kowt/ (a near-homophone of "coat"). One of the goals of speech perception research is to explain how listeners are able to identify these acoustically very different inputs as tokens of the same vowel.

However, the height of the vowel in "caught" is neither random nor a peculiarity of speakers' idiolects. Rather, it is systematically related to the speaker's region of origin, an important dimension of identity. The /kAt/ version is characteristic of the dialect spoken in Boston and environs. Philadelphians are likely to pronounce it /k„t/, and /kowt/ can be heard in inland North Carolina, as well as elsewhere in the South. The adult speech of individuals who learned to speak in one or another of these regions is likely to reflect the regional dialect. While this variability is not linguistically significant (in the sense that it does not affect the understood meaning of the utterance), it may convey information about the speaker, which in turn can affect how the utterance is responded to. Research on speaker perception studies the effects of this kind of variability. Although speech perception *per se* may be of marginal interest to social psychologists, we believe that the phenomena studied in speaker perception research have great potential for yielding insights into a variety of important social psychological processes.

Applying the concepts and methods of speaker perception research to social psychological processes requires consideration of the physical nature of speech. As little as a generation ago, the acoustic analysis of speech demanded considerable in the way of technical skill and instrumentation. To date, the situation has changed markedly.

Modern computing technology has made it possible for anyone with the appropriate software on a PC to edit speech segments with great precision, to perform extremely sophisticated analyses of naturally-produced speech, to quantify subtle variation, to alter speech parameters and resynthesize the speech, etc. This technical capability opens up a world of empirical possibilities that most social psychologists have never contemplated. Obviously, realizing this potential requires some familiarity with the rudiments of speech science: the essentials of the speech production process, the phonological and acoustic structure of speech, and methods of analysis and synthesis. Many psychologists will have encountered some of this material in courses in psycholinguistics, language acquisition, etc. Denes and Pinson (1993) provide a useful introduction to the physics and biology of spoken language; for a somewhat broader survey, see Kent (1997). An excellent overview of phonetic and linguistic markers in speech can be found in Laver and Trudgill (1979).

We will illustrate the potential of this approach with examples from research in two areas, but it must be stressed that these examples are intended to illustrate some of the possibilities of the approach, not to define its limits. The two areas are: (1) social factors affecting within-speaker variability, and (2) effects of interaction on conversational participants' speech.

Social factors affecting within-speaker variability

Even when uttered by the same speaker, a given phoneme or word will vary acoustically on different occasions of articulation. Some of the variability has its origin in linguistic processes, and is not of particular interest to a social psychologist.[2] Even in repetitions of the same sentence, words may differ acoustically,[3] due to a variety of factors. After a brief overview of speech production, we will discuss two kinds of factors that are of particular interest to social psychologists: the speaker's internal state and his/her situated identity.

The mechanisms that underlie vocal production are described by Source-Filter Theory, first proposed in the 19th century by Johannes Müller (1848). According to the theory, air expressed from the lungs causes the vocal folds (formerly called "vocal chords") to vibrate, producing a harmonically rich waveform that is the source of vocal production. The glottal impulses produced by the vibrating vocal folds are filtered by the

supra-laryngeal vocal tract, attenuating some frequencies and amplifying others. By adjusting subglottal air pressure and vocal fold tension, the speaker can vary loudness and pitch. And by configuring the mobile articulators (soft palate, tongue, lips and jaw), the speaker can modify the vocal tract's shape, hence its acoustic filtering characteristics, producing the kinds of variations in sound we identify as speech.

In discussing the vocal expression of emotion, Bacherowski (1999) distinguishes between source-related and filter-related vocal cues. Source-related cues derive primarily from vocal fold vibration, and are reflected in speech as variations (and variability) in pitch and loudness. Filter-related effects reflect configurations of the vocal tract that are sometimes associated with internal states; the same utterance will sound different depending on whether the speaker is smiling or frowning. There is a tendency to think of source-related effects as involuntary and more-or-less direct reflections of autonomic functioning; indeed, fundamental frequency (F0: the rate of vibration of the vocal folds) has been used as a measure in lie detection schemes (Ekman, Friesen & Scherer, 1976; Streeter, Krauss, Geller, Olson, & Apple, 1977). However, the ANS can also produce filter-related cues (e.g., dry mouth, muscle tenseness, etc.) that affect speech. Finally, both source and filter-related effects also have voluntary components.

Identity and Situation. Any listener is capable of identifying the voices of dozens, perhaps scores, of people from brief voice samples with standardized content. Obviously, we are able to do this because people's voices differ acoustically. Differences in the way speakers sound derive from two main sources: individual differences in anatomy that result in acoustic differences in the speech produced, and individual differences in dialect, accent, and speech habits (sometimes referred to as a speaker's *ideolect*). What may be less obvious is that the same sources of variability can provide cues to the aspects of the identity of speakers we *don't* know. For example, a person's age, height, and sex can be judged from his/her voice with surprising accuracy. Krauss, Freyberg and Morsella (2002) found that estimates of these attributes made from a two-sentence voice sample were only marginally less accurate than those made from full-length photographs. The ability to judge a speaker's age from voice is a consequence of physiological changes that accompany aging, and the ability to judge height reflects the correlation of height and laryngeal size (in turn, influencing source characteristics of

7

speech). Yet, the ability to identify a speaker's sex from his or her voice is more complex. In the Krauss et al. study, naïve judges identified 40 speakers as male or female with perfect accuracy. However, although men's and women's voices on average differ on a number of acoustic dimensions, there is no single feature or known subset of features that reliably distinguishes them (Klatt & Klatt, 1990). This leads to the speculation that men and women use their voices differently, and these dynamic differences contribute to our ability to distinguish male and female speakers. There is some evidence, for example, that men and women differ in where within their pitch range they place their voices, with men tending to place their voices in the lower part of their range (Graddol & Swann, 1983).

Perhaps the most widely studied socially significant aspect of voice quality is dialect and accent. A dialect is a variant of a language that is distributed either regionally or by social class. Accent refers to the phonological component of dialect (dialect is also reflected lexically and syntactically), and often is apparent in the speech of people speaking a language that is not their native language.[4] The study of dialects is a central research focus in sociolinguistics, and a great deal is known about their structure, origins, distribution and change (Edwards, 1985; Labov, 1972; 1994; Trudgill, 1983). What makes them of especial interest to social psychologists is their relation to the speaker's identity.

Identity concerns people's sense of who they are—the attributes and features that, on the one hand, distinguish them from others, and, on the other hand, make them members of coherent classes or categories of like individuals. Every person comprises a variety of identities, only a subset of which will be active at any given moment. Social psychologists have tended to focus on the social dimensions of identity (cf. Deaux, 1996), but we will take a somewhat broader view of the concept and distinguish between two general types of identity: *social identity* (defined by the social groups or categories to which he or she belongs, or with which he or she identifies and/or is identified) and *personal identity* (socially relevant aspect of the individual's physical and psychological make-up). Many attributes of these aspects of identity are embodied in speech.

In addition to these relatively general and enduring features of individuals' speech, it is also the case that the same individual will speak differently on different

occasions. One source of this variability is captured by the sociolinguistic concept of *register*. A speech register is a type of situated linguistic variation that is conditioned by occasions of usage. Registers differ from dialects, which are variations conditioned by region or social status (Ferguson, 1983). Registers can constitute very broad categories of usage (e.g., casual vs. formal speech) or more narrowly formulated varieties such as "motherese" or "sports-talk" (the speech style affected by sportscasters). Since registers vary with situation, the register a speaker employs directly reflects his or her definition of the situation, the social role he or she is playing in the situation, or the identity that is active in the speaking situation.

Registers affect all levels of linguistic analysis—lexical, syntactic and phonological. The formal register, for example, eschews many of the colloquial terms and syntactic forms often found in casual speech. The formal "Ask him if he wants a soft drink" might be rendered "Ask him do he want a pop" in an informal setting. At the articulatory level, casual speech is typically marked by *reduction*—the tendency to articulate unstressed consonants imprecisely, and to diminish distinctions among vowels by centralizing them. In a phrase like "going fishing," the vowel sequence in the first word might be elided to /o/ and the nasalized /ing/ suffixes might lose their /g/s, so that "going" rhymes with "bone."

It is important to acknowledge the range of within-speaker variability. Although a dialect may be an enduring characteristic of an individual's speech, there is considerable variation from occasion to occasion in how distinctly the dialect is rendered. Even the stereotypical working-class New Yorker who typically omits /r/ after vowels ("fourth floor" → "fawth flaw") will articulate them quite clearly in formal or emphatic speech (Labov, 1966). Variation in register is even more apparent. For social psychologists, this variability has important significance. Because dialect and register are associated with speakers' identities and definitions of the situation, in effect, the way a speaker speaks reflects the identity (or dimension of identity) that is active and the way he/she defines the interaction situation.

Internal state. A speaker's speech will vary qualitatively from occasion to occasion depending on a variety of internal state factors, and even as small a speech sample as "Hello" can contain enough information for a familiar listener to determine

whether the speaker is excited, depressed, annoyed, etc. Several investigators have examined the properties that distinguish different emotions (Fairbanks; Williams; Scherer), but because for practical reasons these analyses have been based largely on actors' portrayals of emotions rather than naturally-occurring expressions, it's difficult to reach any firm conclusions about the acoustic correlates of emotional speech. It seems fairly clear that such intense emotional states as anger, fear and joy are associated with elevated F0, reflecting the physiological arousal that accompanies these states. These findings are corroborated by studies using natural speech samples. The voices of speakers experiencing high levels of stress are characterized by elevated F0 (Williams & Stevens, 1969; Streeter, et al., 1983). F0 is a useful index because it is relatively easy to compute even under less-than-ideal acoustic circumstances, but it reflects only one aspect of voice. Efforts to find vocal indices of other aspects of emotion have been generally less successful.

All of the foregoing would be of some interest to social psychologists even if listeners were not attentive to variability in speakers' voices. However, there is considerable evidence that naïve listeners respond quite sensitively to speech variation, and that these perceptions have important evaluative and attributional effects. For example, social psychologists (among others) have studied the effects on listeners of deviations from standard or prestige dialect. Generally speaking, speech patterns associated with stigmatized or socially devalued identities elicit negative evaluations of the speaker and the speech's contents (Callan, Gallois, & Forbes, 1983; Genesee & Bourhis, 1988; Giles, Taylor, & Bourhis, 1977; Lambert, Hodgson, Gardner, & Fillenbaum, 1960; Ryan, Bourhis, & Knops, 1991; Ryan & Capadano, 1978).

Modern speech processing technology has made it relatively easy to alter specific parameters of natural speech and resynthesize the speech, allowing investigators to study the effects of these changes on listeners. For example, Brown, Strong and Renscher (1974) and Apple, Streeter and Krauss (1979), using an analysis/resynthesis procedure, found that elevating speakers' F0s caused them to be perceived as weaker, less benevolent, competent, truthful and persuasive, and more nervous. Because the analysis/resynthesis procedures available at the time these studies were carried out did a poor job of reproducing women's voices, both studies used male speakers exclusively,

and their results probably should not be generalized to women. As noted above, there is some reason to suspect that pitch placement differs for men and women (Graddol & Swann, 1983), and a preliminary study in our lab suggests that the effects of elevated F0 on listeners may vary for male and female speakers (Gardner, 2003).

Effects of social interaction on speech

Most accounts of language development assume that language is more-or-less fixed by the onset of puberty, yet changes can occur in an adult speaker's phonetic repertoire. The changes may be subtle—it can extremely difficult to acquire a new language or to eliminate an accent—but speakers well beyond the critical period of language development adjust their speech during conversation. Moreover, at least some of these changes seem to have social significance. Two recent psycholinguistic studies have examined the potential for phonetic convergence indirectly, one by measuring the end results of long periods of time in different linguistic environments (Sancier & Fowler, 1997), and the other by measuring the effect of immediately repeating another person's speech (Goldinger, 1998). In both studies, speakers adjusted their speech enough that the change was both acoustically measurable and discriminable by listeners: Sancier and Fowler found changes in a bilingual speaker's L1 speech after she spent time in an L2-speaking environment. Goldinger found that listeners judged modeled repetitions to be better imitations of a model's productions than earlier productions of the same words. These studies hint at a process of linguistic modification that ought to occur in interacting dyads, but this possibility is only beginning to be examined directly.

Social interaction is the most common and natural setting for language use, and language is a critical element in most social interaction. The study of language use in social interaction takes many forms; here we will focus on changes in speech that occur as a function of conversational interaction. Currently, there are three prominent psychological approaches to language in social interaction: the research programs identified with Howard Giles, Herbert Clark, and Simon Garrod, and their colleagues. Common to all three approaches is an emphasis on social factors in dialogue (as opposed to individual factors in speech production) as the basic unit of analysis for language. From this perspective, language, as it is used in social interaction, consists of more than

exchanges of grammatically well-formed stings generated by isolated individuals who can be regarded as "autonomous information processors" (Clark & Brennan, 1991). Rather, it is what sociologists call a joint activity that is used to project social categories (Giles, Coupland, & Coupland, 1991; Giles & Powesland, 1975; Giles, Scherer, & Taylor, 1979), to accomplish mutual goals (Clark, 1996; Clark & Wilkes-Gibbs, 1986), or to align representations and mental models (Garrod & Doherty, 1994; Pickering & Garrod, in press).
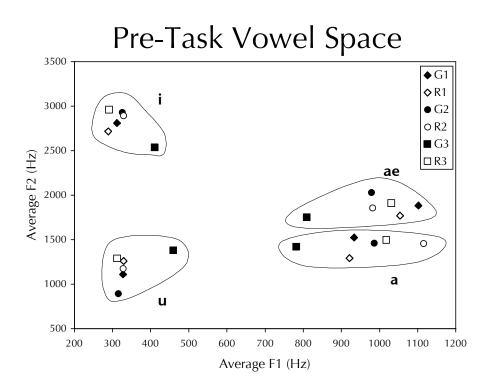
One dialogic function identified by all of these approaches is an increase in similarity in some aspect of participants' linguistic production, variously termed convergence, common ground, or alignment. Interlocutors are known to converge in speech rate (Giles et al., 1991), sub-vocal frequency (Gregory, 1990), and vocal intensity (Natale, 1975); to establish and increase common ground to the exclusion of over-hearers (Schober & Clark, 1989); and to align description schemes (Garrod & Doherty, 1994) and syntactic constructions (Branigan, Pickering, & Cleland, 2000).

Although studying language in an interactive setting is an important improvement over standard psycholinguistic practice, all three approaches pay scant attention to variation in the most basic level of analysis in speech—the phonological level. In speech perception, a listener identifies a phonetic token as a member of a phonological category. Most phonetic categories are distinguishable by variation in the two lowest frequency components of speech (produced by the vocal tract filter), which are called the first and second formants (F1 and F2). But the perception process preserves some distinctions between tokens of the same category. This phenomenon is most clearly seen in vowels, as in the difference between regional dialectal pronunciations of the word, "pen." For U. S. Midwesterners, the vowel in "pen" is almost identical to the vowel in "pin," while New Englanders maintain the distinction in vowel color. Thus, the phoneme category for /E/ includes a broad range of variants.[5] Interactive convergence has been found for F0 (heard as overall pitch), speaking rate, and amplitude, but none of these acoustic parameters differentiate phonological categories in English as F1 and F2 do, and since the phonological dimension of speech carries special social significance, interactive convergence in this acoustic dimension can be especially informative.

An interesting study by Bourhis and Giles (1977) illustrates how this approach might be applied. Bourhis and Giles had judges assess degree of accentedness in Welsh speakers who had overheard an English experimenter making negative comments about the Welsh language. The Welsh accents of participants with strong Welsh identities became stronger following this challenge to their identities. However, it is unclear exactly what acoustic-phonetic attribute contributed to the measure of accentedness in this study. Ideally, one would like to represent the response to the identity challenge by a measure that is objective and quantitative—e.g., changes in vowel color as indexed by their F1 and F2 frequencies.

To demonstrate the use of this approach, we present some preliminary data in ongoing research using the HCRC Map Task (Anderson et al., 1991). The Map Task is a structured dyadic task in which the participants use maps with labeled landmarks and try to reproduce on one map the path to a destination that appears on the other through purely conversational interaction. The task stimulates conversational interaction with a great deal of between-speaker lexical repetition, which is necessary for examining phonetic convergence. In addition, the task permits assignment of different social roles; one member of a pair (the "instruction giver") has maps with paths that must be duplicated by the other member (the "instruction receiver"), creating a role asymmetry.

Pardo (2001), using an AXB psychophysical judgment procedure, found that participants' speech became more similar.[6] The initially divergent pronunciations of landmark names converged as a result of interaction. What may be more interesting is a possibility revealed by acoustic measures of vowels produced before and after the interaction. These preliminary analyses suggest that speakers' global production, not just their pronunciations of specific site names, were affected as a result of interaction (Pardo, 2003). Moreover, global pronunciation may have been influenced by the speaker's role. Based on the data analyzed thus far, instruction givers' vowels centralized more than instruction receivers' vowels. The data are shown below in Figure 1, which plots the vowel spaces for three female dyads before (1A) and after (1B) performing the Map Task. Thus, what has formerly been regarded as random variation in speech production is now linked to the social dynamics of conversational interaction. Making this connection is only the beginning of a promising avenue of research.

A.

Pre-Task Vowel Space
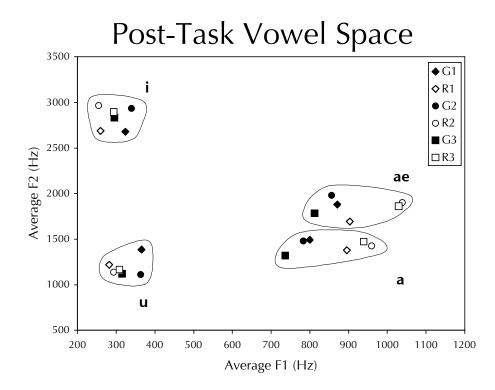

B.

Post-Task Vowel Space

Figure 1: Comparison of acoustic vowel spaces for vowels produced before (A) and after (B) participation in a paired conversational task by 3 pairs of female talkers. Each graph plots average F1 by F2 values measured at the midpoint of each of three vowel tokens spoken in hVt words (/i/ in heat, /ae/ in hat, /a/ in hot, /u/ in hoot). Filled bullets represent items spoken by Givers in the task and open bullets represent items spoken by Receivers in the task. Bullets of the same shape represent talkers who were paired together in the task. Note the overlap in vowel spaces for Givers and Receivers pre-task and the centralization of Givers' tokens post-task, particularly for /ae/ and /a/.

Concluding Comment

Social psychologists' growing appreciation of the role that language plays in social life has focused their attention on the extent to which the forces that shape our social behavior are mediated by, and reflected in, the words we utter. This realization, and the research that it stimulated, has yielded important insights into some of the field's core phenomena. In this essay, we have tried to make the case that investigations utilizing the sound structure of speech can contribute new and different insights. Although we have described several research efforts utilizing this approach, we expect that social psychologists will employ it to pose new and exciting questions that we have not ventured to ask here.

Ironically, although sociolinguists associate speech variation with social categories (e.g., Labov, 1974), psycholinguists have expended considerable effort to eliminate its effect on categorization (see Johnson & Mullenix, 1997). For a psychologist interested in social interaction or language use, this unclaimed variability is potentially an empirical treasure trove. The underlying theoretical question in speech perception research derives from the fact that speech is acoustically variable in its realization, both between and within speakers, despite marked consistency in perception. Somehow, a listener is able to overcome the acoustically disparate productions of phonemes to arrive at what each speaker intended to say. Yet, after over 70 years of intensive research in psycholinguistics, the details of this process still elude full explanation. Because much of the variability in speech can be traced to non-linguistic factors that have social significance, it would not be surprising if investigations of speech by social psychologists yielded important insights into the role variability plays in speech perception. Likewise, speech research offers a new arena for research on social behavior.

References

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S., & Weinert, R. (1991). The HCRC Map Task corpus. *Language & Speech*, *34*, 351-366.

Apple, W., L. A. Streeter, & Krauss, R.M. (1979). Effects of pitch and speech rate on personal attributions. *Journal of Personality and Social Psychology,* 37, 715-727.

Bachorowski, J.-A. (1999) Vocal expression and perception of emotion. *Current Directions in Psychological Science,* 8, 53-57.

Bourhis, R. Y. & Giles, H. (1977). The language of intergroup distinctiveness. In H. Giles (Ed.), *Language, Ethnicity, and Intergroup Relations*. London.

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition, 75,* B13-25.

Brown, R. and A. Gilman (1960). The pronouns of power and solidarity. In T.A. Sebeok, (ed.), *Style in Language*.. Cambridge, MA, MIT Press. .

Clark, H. H. & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22,* 1-39.

Clark, H. H. (1996). *Using Language*. Cambridge University Press.

Denes, P.B., & Pinson, E.N. (1993). *The speech chain.* New York: W.H. Freeman & Co.

Ekman, P., Friesen, W.V., Scherer, K.R. (1976). Body movement and voice pitch in deceptive interaction. *Semiotica*, 16, 23-27.

Gardner, R.C. (2003). *What's in a voice? Pitch and persuasiveness.* Unpublished Honors Thesis, Department of Psychology, Columbia University

Garrod, S. & Doherty, G. (1994). Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition*, *53*, 181-215.

Giles, H. & Powesland, P. F. (1975). *Speech styles and social evaluation*. Academic Press.

Giles, H., Coupland, J., & Coupland, N. (1991). *Contexts of Accommodation: Developments in Applied Sociolinguistics*. Cambridge University Press.

Giles, H., Scherer, K. R., & Taylor, D. M. (1979). Speech markers in social interaction. In K. R. Scherer & H. Giles (Eds.), *Social Markers in Speech* (pp. 343-381). New York: Cambridge University Press.

Giles, H., Taylor, D.M., & Bourhis, R.Y. (1977). Dimensions of Welsh identity. *European Journal of Social Psychology*, 7, 165-174.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105,* 251-279.

Gradol, D. and J. Swann (1983). "Speaking fundamental frequency: Some physical and social correlates." Language and Speech, **26**: 351-366.

Gregory, S. W. (1990). Analysis of fundamental frequency reveals covariation in interview partners' speech. *Journal of Nonverbal Behavior, 14*, 237-251.

Holtgraves, T. and J.-N. Yang (1990). Politeness as universal: cross-cultural perceptions of request strategies and inferences based on their use. *Journal of Personality and Social Psychology,* 59, 719-729.

Johnson, K. & Mullenix, J. W., Eds. (1977). *Talker Variability in Speech Processing*. San Diego: Academic Press.

Klatt, D.H., & Klatt, L.C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America*, 87, 820-857.

Labov, W. (1966). *The Social Stratification of English in New York City.* Washington, D.C.: Center for Applied Linguistics, 1966.

Labov, W. (1974). Linguistic change as a form of communication. In A. Silverstein (Ed.), *Human Communication: Theoretical Explorations* (pp. 221-256). Hillsdale, NJ: Lawrence Erlbaum Associates.

Lambert, W,E., Hodgson, R.C., Gardner, R.C., & Fillenbaum, S. (1960). Evaluational reactions to spoken language. *Journal of Abnormal and Social Psychology,* 60, 44-51.

Müller, J. (1848). *The physiology of the senses, voice and muscular motion with the mental faculties.* London: W. Baly, Walton & Maberly.

Natale, M. (1975). Convergence of mean vocal intensity in dyadic communication as a
function of social desirability. *Journal of Personality and Social Psychology, 32,*
790-804.

Pardo, J. S. (2001). Imitation and coordination in spoken communication. *Dissertation
Abstracts International: Section B: the Sciences & Engineering. Vol 61(10-B),
May 2001, 5576, US: Univ Microfilms International.*

Pardo, J. S. (2003). On phonetic convergence and conversational interaction.
Manuscript in preparation.

Pickering, M. J. & Garrod, S. (in press). Toward a mechanistic psychology of dialogue.
Behavioral & Brain Sciences.

Sancier, M. L. & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian
Portuguese and English. *Journal of Phonetics, 25*, 421-436.

Schober, M. F. & Clark, H. H. (1989). Understanding by addressees and overhearers.
Cognitive Psychology, *21*, 211-232.

Semin, G. R. and K. Fiedler (1991). The linguistic category model, its bases, applications
and range. IN, W. Stroebe and M. Hewstone (Eds.), *European Review of Social
Psychology,* II. London, John Wiley & Sons.

Streeter, L. A., Macdonald, N. H., Apple, W., Krauss, R. M. & Galotti, K. M. (1983).
Acoustic and perceptual indicators of emotional stress. *Journal of the Acoustical
Society of America*, **73**,1354-1360.

Streeter, L. A., Macdonald, N. H., Apple, W., Krauss, R. M. & Galotti, K. M. (1983).
Acoustic and perceptual indicators of emotional stress. *Journal of the Acoustical
Society of America*, **73**,1354-1360.

Streeter. L. A., Krauss, R. M., Geller, V., Olson, C. T., & Apple, W. (1977). Pitch
changes during attempted deception. *Journal of Personality and Social
Psychology*, 35, 345-350.

Williams, C. E. and K. N. Stevens (1969). On determining the emotional state of pilots
during flight: An exploratory study. *Journal of the Acoustical Society of America,*
40, 1369-1372.

Williams, C. E. and K. N. Stevens (1972). Emotions and speech: Some acoustical
correlates. *Journal of the Acoustical Society of America,* 52, 233-248.

[1] Writing and sign are two other forms in which language is made tangible. Word-syllabic writing systems are essentially nonphonetic transcriptions (English is a bad example) that record mainly the semantic content of speech. Sign systems are full-fledged languages, but they lie beyond the scope of this chapter.

[2] Three examples: (1) To produce the English phoneme /k/, the rear part of the tongue's surface (the *dorsum)* is raised to briefly contact the soft palate (*velum*). However, the initial /k/ sound in the words "keep" and "cool" are not articulated identically. In "keep" the dorso-velar contact is further forward than it is in "cool," in anticipation of the following vowel—a process called *coarticulation.* (2) When an English speaker articulates a word like "pop," the initial and terminal /p/s will differ acoustically because of a rule of English phonology that distinguishes terminal and nonterminal aspirants. (3) The word "John" will be articulated differently in the sentences "The window was broken by John" and "John broke the window" because of differences in the two sentences' prosodies.

[3] It is probably the case that no two utterances of the same token are *ever* identical acoustically, although the differences may be too small for even a trained listener to apprehend.

[4] As a rule, languages acquired after adolescence are marked by the phonetic structure of the speaker's first language. Typically phonemes in L2 that don't occur in L1 will be assimilated to similar L2 phonemes, as when native German speakers say "ze" for the English word "the".

[5] The same is true for consonants, but the differences are less salient.

[6] Following Goldinger (1998), the AXB task takes an item produced by one member of a pair (X) and surrounds it by two repetitions of the item spoken by the other member (A & B)—the item that was a conversational repetition (A/B) and a version that was produced before the conversation (B/A). Ordinary listeners choose which item, A or B, sounded more like the middle item, X, in pronunciation.