RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 27 (2005)
Indiana University

Indexical and Linguistic Channels in Speech Perception: Some Effects of Voiceovers on Advertising Outcomes [1]

Susannah V. Levi and David B. Pisoni

Speech Research Laboratory
Department of Psychological and Brain Sciences
Indiana University
Bloomington, Indiana 47405

[1] Preparation of this chapter was supported by grants from the National Institutes of Health to Indiana University (NIH-NIDCD T32 Training Grant DC-00012 and NIH-NIDCD Research Grant R01 DC-00111). We wish to thank Luis Hernandez and Darla Sallee for technical assistance and help with this manuscript.

Indexical and Linguistic Channels in Speech Perception: Some Effects of Voiceovers on Advertising Outcomes

Abstract. This article examines the effects that voice features have on advertising. Previous research in neurolinguistics and psycholinguistics shows that linguistic and extralinguistic ("indexical") properties of speech are closely coupled in speech perception and spoken language processing. We review research from the advertising and marketing literature that examines which voices are the most suitable for voiceovers, whether speech rate compression is advisable, and under what circumstances voice selection is most important. We integrate these two bodies of literature and conclude that the voices used in advertising should be familiar and consistent across the campaign and the speaking rate may be increased without deleterious effects.

Components of Speech

Marshall McLuhan wrote "the medium is the message." That is, not only is the content of the message itself important in conveying information, but so too is the medium, or the way in which the intended message is conveyed to an audience. When people perceive spoken language, information about the content of the message is transmitted to the listener, along with information about the specific person who produced the message. Because these two sources of information are ineluctably bound together in the speech stream, both channels of information contribute to the final product of perception and both should be considered by advertisers when developing voiceovers.

Speech is a complex, multimodal time-varying pattern. Although both auditory and visual cues function in speech perception, we will focus only on the auditory portion. Spoken language encodes two different sources of information. First, it carries linguistic information about the symbolic content of the talker's intended message. This content contains several levels of linguistic information: phonological (sounds), morphological (units which form words), syntactic (combining words into sentences), and semantic (meaning of an utterance). Taken together, this linguistic information provides the content of an utterance.

The second type of information that is carried in the speech stream is often termed paralinguistic, extralinguistic, or indexical. Indexical information can be thought of as the "medium" through which the message is conveyed. Abercrombie (1967) wrote that "[s]uch 'extra-linguistic' properties of the medium… may fulfill other functions which may sometimes even be more important than linguistic communication, and which can never be completely ignored" (p. 5). Abercrombie divided the indexical properties of speech into three sets: (1) those properties that indicate group membership (e.g., regional, dialectal, and social aspects of speech), (2) those that characterize the individual (e.g., age, gender, and size and shape of the vocal tract), and (3) those that reveal changing states of the speaker (e.g., affective properties such as fatigue, excitement, amusement, anger, suspicion, health, speaking rate). Indexical and linguistic information in speech correspond to what cognitive psychologists often refer to as source and item information, respectively (see Hilford, Glanzer, Kim, & DeCarlo, 2002).

What makes speech a complex signal is that these two properties are carried simultaneously in a single acoustic waveform that is at first produced by an individual speaker and then perceived by a listener who can extract both sources of information. Speech is generated by a speaker's larynx and supralaryngeal vocal tract. The vocal tract which extends from the larynx through the throat and mouth to

the lips acts as an acoustic filter, enhancing certain resonance frequencies (formants) and attenuating others. When speakers produce different sounds in a language, they constrict their vocal tract at different locations. Which frequencies are enhanced or attenuated in the vocal tract is determined both by its length and by the location of the constriction. In the productions of sounds, the relative frequencies provide the linguistic information about the place of constriction of sounds. In contrast, the absolute frequencies that resonate in a particular person's vocal tract are dependent on the length of that person's vocal tract and thus provide talker-specific information. The sound spectrogram in Figure 1 provides a specific example of the integration of linguistic and indexical properties of speech in the production of speech. The formant values produced by the female speaker (first author) are higher than those produced by the male speaker (second author), showing one indexical difference resulting from differences in vocal tract length. The overall movement and relative locations of the formants, on the other hand, provide linguistic information and indicate that the speakers are saying the same utterance. Thus, the same vocal mechanisms produce both linguistic and indexical information simultaneously and both sources of information are encoded and carried in the same signal.

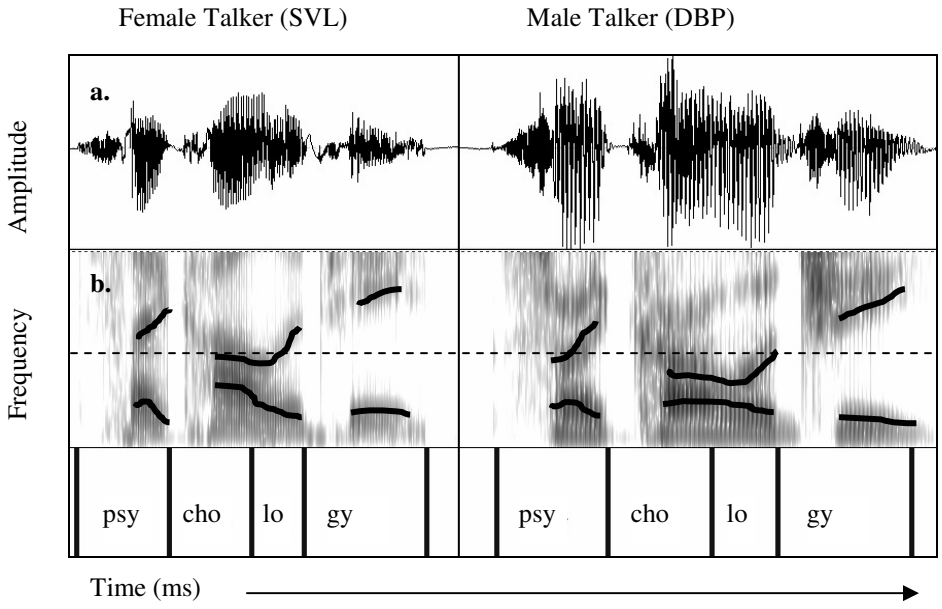

Figure 1. Waveform (a) and spectrogram (b) of the word "psychology" produced by the first author (SVL) and the second author (DBP). Dark lines in the spectrogram represent the first formant (lower curve) and second formant (upper curve).

The perception of these two different aspects of speech is illustrated in Figure 2. The basilar membrane (bottom of Figure 2) is situated in the cochlea in the inner ear and allows a listener to segregate frequencies. The left path in Figure 1 shows the absolute frequencies that are heard by the listener and provide indexical information about an individual talker. The right path represents the relative frequencies which provide linguistic information about the intended message. In this way, both the linguistic and the indexical properties of the speech signal can be perceived and encoded by the listener.

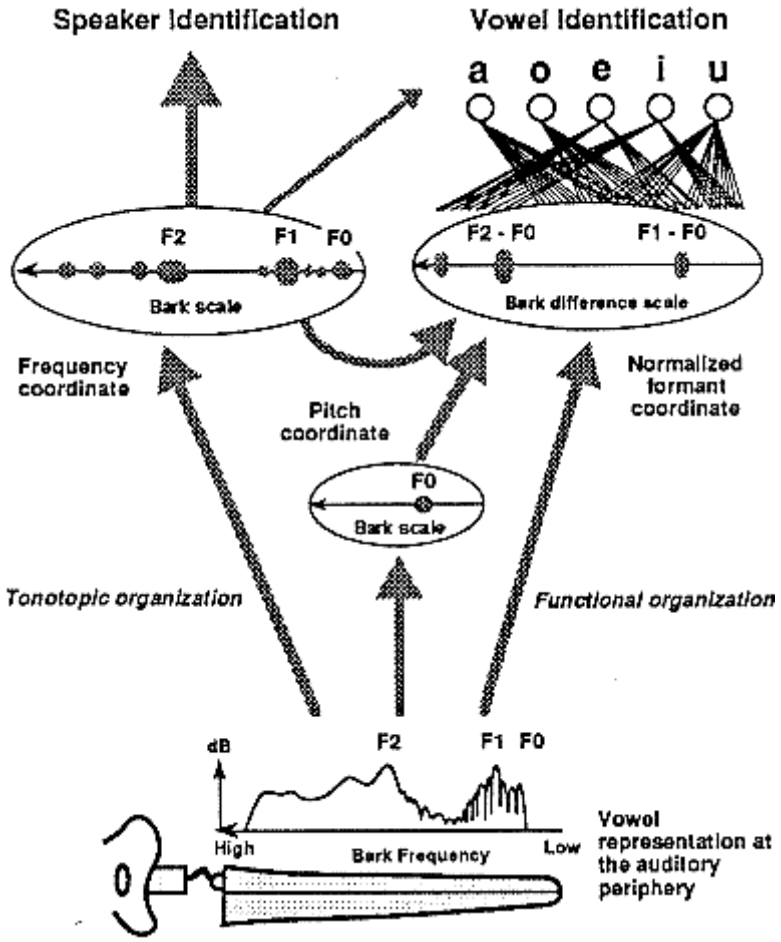

Figure 2. Representation of auditory perception of both indexical and linguistic properties of speech. The absolute frequencies (left side) provide speaker identification, while the relative frequencies (right side) provide vowel identification (Hirahara & Kato, 1992).

The remainder of this chapter is organized into three sections. In the first, we review several lines of neurolinguistic and psycholinguistic research on the perception of indexical properties of speech. The findings discussed in this section confirm that two distinct channels of information are carried in the speech signal. Moreover, the results suggest that the processing of one set of properties affects the processing of the other. In the second section, we consider research from the advertising and marketing literature that examines which voices are the most suitable for voiceovers, whether speech rate compression is advisable, and in what contexts selecting the appropriate voice is most important. In the last section, we integrate these two separate bodies of literature in order to determine what kinds of voices should be used for the most effective advertising.

The Science of Voice Processing

Behavioral and neural studies on the perceptual processing of speech illustrate its bipartite nature. By asking listeners to attend to either the linguistic or the indexical (voice) properties of speech, neuroscientists have shown that these two aspects of speech are processed differently in the brain.

Despite this difference in neural processing, behavioral studies show that the two are in fact closely linked and that voice (indexical) characteristics affect linguistic processing of speech.

Neural Processing of Voices

Neural studies of voice identification and discrimination reveal that characteristics of the voice are processed in brain areas which are distinct from those that process the linguistic properties of the speech signal. In an early study of hemispheric specialization, Landis, Buttet, Assal, and Graves (1982) found that while both hemispheres can be utilized in voice recognition, there was a distinct advantage of the left hemisphere for linguistic tasks. Landis et al. played monosyllabic consonant-vowel words into either the right or the left ear. In the linguistic task, listeners were asked to press a button every time they heard a specific target word. Listeners' reaction times showed a clear right-ear advantage (REA), responding faster when the target word was presented to the right ear than the left. Because the two hemispheres control contralateral body functions, showing a preference for the right ear, indicates that the left hemisphere dominates in the linguistic task. In a second experiment, listeners were asked to push a button when they heard a particular male or female voice. In this study, female voices elicited a REA, but male voices a left-ear advantage (LEA). Landis et al. interpreted these results by remarking that higher frequencies have been shown to elicit a REA and that female voices, with their higher pitch and formants, may therefore also be processed with a REA. The major finding of this study was the demonstration that both hemispheres are involved in voice recognition, whereas word recognition displays left hemisphere dominance.

Kreiman and Van Lancker (1988) found similar results using a dichotic listening paradigm. In a dichotic listening task, listeners hear different words presented simultaneously in both ears and are asked to attend only to the stimuli that are played in either the right or the left ear. Using a set of 50 famous male voices, they asked listeners to write down both the word (linguistic task) and the person who said the word (indexical task). As expected, they found a clear REA in the word recognition task. The results of the voice identification task were less conclusive. Listeners showed no ear advantage for the voice recognition task, consistent with the earlier results of Landis et al. (1982). They did, however, find a *relative* left-ear advantage; that is, relative to the word recognition task, listeners showed a greater advantage for the left ear.

More recent studies have been able to isolate voice processing to more specific brain regions. Glisky, Polster, and Routhieaux (1995) tested elderly listeners' ability to recall either the content or the voice of previously heard sentences. They found that listeners with high frontal lobe function outperformed those with poor frontal lobe function on the voice task, but showed no difference in their performance on the sentence recall task. Conversely, listeners with high medial temporal lobe function outperformed listeners with low function in the sentence recall task, but did not differ on the voice task. These results confirm that the processing of voice information is independent of linguistic processing.

More recently, using functional magnetic resonance imaging (fMRI), Stevens (2004) reported distinct brain regions for voice- and word-discrimination tasks. Listeners were asked to determine whether two talkers were the same or whether two words were the same. Stevens found that attending to either the word or the voice altered the functional activity of the brain. In particular, the voice comparison task produced activation in the right fronto-parietal area, whereas lexical processing was associated with increased activation in the left frontal and bilateral parietal areas.

Other studies have shown that voice processing can be further subdivided; familiar voices are processed differently than unfamiliar voices. In these studies, familiarity refers to people who were

personally known to the listeners.[2] Using fMRI, Shah et al. (2001) found that familiarity of voices and faces resulted in increased activity in the posterior cingulate cortex as compared to unfamiliar voice and face processing. Nakamura et al. (2001) also found different brain areas involved in familiar versus unknown voice processing using positron emission tomography (PET).

Taken together, these studies of the neural processing of speech demonstrate that the indexical (source) properties are indeed distinct from the linguistic (symbolic) properties of speech, despite the fact that they are carried simultaneously in the same speech waveform. When listeners are asked to attend to voice characteristics of the speaker, they utilize different areas of the brain than when they process the linguistic information in the signal.

Interactions of Indexical (voice) and Linguistic Processing

Although the studies reviewed in the previous section revealed that distinct brain areas are involved in voice perception and linguistic processing, results of behavioral studies indicate that these two properties of the speech signal are closely coupled functionally. Properties of the voice affect the processing of linguistic information. Most important for the concerns of advertisers is the incidental or indirect effects of voice information on the processing of the content of the message which show that consistency and familiarity of the voice facilitates linguistic processing of the message.

Evidence from a variety of behavioral studies shows that consistency of the voice is an important aspect of linguistic processing. Using a speeded classification task, Mullennix and Pisoni (1990) asked listeners to categorize a set of spoken words that differed on two perceptual dimensions: the linguistic dimension in which the initial sound of the words varied between “p” and “b” and the indexical/gender dimension in which words were spoken by either a male or a female talker. In the control conditions, a single talker produced all words, thereby holding the indexical dimension constant. In the orthogonal conditions, the two dimensions varied randomly so there was no consistency between the two dimensions. Listeners were asked to classify words using each dimension separately, ignoring possible variation along the other dimension. Mullennix and Pisoni found that reaction times were slower in the orthogonal conditions than in the control conditions, indicating that listeners were not able to “filter out” the indexical variation while performing the linguistic task and that variation of the voice inhibits listeners’ performance. They also found that increasing the number of talkers from two to 16 had an even greater effect of slowing down classification times. This study revealed that the indexical properties of speech are not processed independently of linguistic content of the signal and that irrelevant variation in a non-attended perceptual dimension (in this case, the indexical dimension) is not discarded when performing such a task, but is instead processed in an integral manner.

Schacter and Church (1992) found a similar same-voice advantage in a stem completion task. In the study phase, listeners heard a series of words and rated either the pleasantness of the word or the pitch of the voice. In the test phase, listeners heard a series of syllables mixed with noise and were asked to write down the first word that came to mind. Schacter and Church found that when voices of the study words and the test syllables matched, a greater priming effect was observed than when the voices were switched. In other words, listeners were more likely to respond with a word they had heard during the

[2] In a several studies, Van Lancker, Kreiman, and colleagues (Van Lancker & Kreiman, 1987; Van Lancker, Cummings, Kreiman, & Dobkin, 1988; Van Lancker, Kreiman, & Cummings, 1989) showed that recognizing famous voices and discriminating between unfamiliar voices engaged different brain areas. It is not possible to conclude from these studies that famous and unfamiliar voices themselves are processed differently because the two tasks were fundamentally different. In the famous voice recognition task, listeners were asked to name the famous voice and to draw on long term memory. In the unknown voice discrimination task, listeners compared two unknown voices that were presented one following the other.

study phase of the experiment if that studied word was spoken in the same voice as the syllable heard during the test phase.

In another study, Goldinger (1996) showed that listeners exhibit a same-voice advantage in recognition memory when performing a linguistic task. Listeners were asked to type the word they heard when it was presented in noise. Test words that were spoken by the same talker were recognized more often than words spoken by a different talker. Perhaps even more striking was the finding that the same-voice advantage did not decline significantly across different delays between study and test. Listeners who returned after a week showed the same voice advantage as those who returned after only a five-minute delay, indicating that listeners encode and store information about a voice for an extended period of time, even when the demands of the task do not consciously ask listeners do so. The lack of an effect of delay suggests that the voice effect does not disappear rapidly but is available and stored in memory for an extended period of time. In a separate voice-recognition task, Goldinger (1996) found that listeners' ability to explicitly remember voices did decline with an increased delay. Together these two sets of results suggest that while listeners may lose their ability to explicitly remember the voice, attributes of a voice remain in memory and have effects on language processing for an extended period of time.

Using a list recall task, Goldinger, Pisoni, and Logan (1991) also found an advantage for voice consistency in learning and memory. In this study, listeners first heard 10 words and were subsequently asked to recall the list. Goldinger et al. varied the number of voices which were used to present the list of words and the rate at which the stimuli were presented. The authors found that at fast presentation rates, lists of words produced by multiple talkers were recalled less accurately than lists that were spoken by only a single talker. In contrast, at slow presentation rates, lists produced by multiple talkers were actually remembered more accurately than single-talker lists. Lightfoot (1989) conducted a follow-up study using this same methodology. The difference in Lightfoot's study was that lists were spoken by voices that were familiar to the listeners. Interestingly, voice familiarity caused the advantage of voice consistency to disappear.

In a continuous recognition memory experiment using spoken words, Palmeri, Goldinger, and Pisoni (1993) played long lists of words to listeners and asked them to determine whether each word was "old" (one that had been previously heard) or "new" (one that had not been previously heard). In order to assess the effects of voice on recognition memory, half of the old words were repeated in the same voice and half were repeated in a different voice. As in the previous studies, listeners responded more quickly and more accurately when old words were repeated in the same voice. Palmeri et al. also found that the lag (i.e. the number of words intervening between the first and second presentation of a word) did not interact with the same-voice advantage, indicating that the facilatory effect of maintaining the same voice is robust over time.

In addition to consistency of voices, familiarity with voices facilitates recall and recognition of spoken language. Several studies have shown that familiarity with a set of talkers allows for faster and more accurate linguistic processing. For example, Nygaard, Sommers, and Pisoni (1994) trained listeners to identify ten unfamiliar talkers by name over a period of ten days. During the test phase on the last day, listeners were presented with novel words mixed in noise that were spoken either by the now familiar talkers or by unknown talkers. Subjects were simply asked to identify words and were not required to respond to the voice of the talker. The results indicated that listeners identified novel words in noise better when the words were spoken by familiar talkers, than when the words were spoken by unfamiliar talkers. In a follow-up study, Nygaard and Pisoni (1998) showed that the advantage of talker familiarity extends to sentence-length utterances as well.

The behavioral studies reviewed in this section suggest that the linguistic and indexical channels of speech are closely coupled. In linguistic tasks (e.g. word recognition and phoneme discrimination) that on the surface do not appear to rely on indexical or voice properties, a strong effect of voice is reliably observed. Both familiarity with the voice and consistency of the voice facilitate processing of the linguistic (symbolic) content of the message.

Advertising/Marketing

Advertising messages using spoken language contain meaningful information (the intended message text), visual information (in the case of television advertising), and voice information. It has been shown that when both audio and visual information are present, the auditory information has attentional priority over the visual modality and can mask otherwise distracting information in the visual signal. Drew and Cadwell (1985) varied the angle and zoom of jump cuts in an informational video. They found that when an audio signal accompanied the video there were no negative effects on viewers' attitudes towards the video, showing the importance of an audio stream for maintaining coherence and sufficiently masking distracting visual cues. Since audio information is clearly relevant in both radio and television advertising, selecting an appropriate voice to accompany the product of the advertisement is important and may have significant effects on a wide range of outcome measures. In this section, we consider some factors that are relevant for selecting an appropriate voice for an advertising campaign. We will also discuss under what conditions voice characteristics are likely to affect listeners' attitude towards and memory for the product.

Picking the Right Voice

In selecting the right voice to accompany an ad, several considerations must be made. For instance, should the voice of a famous person be used? What gender voice is appropriate for a given product? Does the accent or nativeness of the talker's voice play a role in listeners' understanding, attitude, and memory for the product?

A first question that an advertiser might consider is whether the spokesperson for a product should be famous. It may be the case that famous actors are better able to read the script of an ad (Alsop, 1987). Not surprisingly, it is also important that if a celebrity is used in an ad, that he/she match the product in such a way that credibility of the product is enhanced (Plapler, 1974; Misra & Beatty, 1990).

While using a celebrity voice in advertising is more expensive, it may be the case that a celebrity is actually better at selling a product than an unknown person. Leung and Kee (1999) conducted an experiment to test whether celebrity spokespeople were better than unknown actors in selling a product. They took a recent television commercial which used two well-known DJs in Hong Kong as the voiceovers for the ad and recorded the same ad with two trained but not well-known actors. Viewers who saw the ad with the celebrity voiceover had higher brand recall and encoded more product brand information, although there was no significant difference in viewers' intent to buy the product.

Finding the right talker for a voiceover also includes deciding the appropriate gender of the speaker. While male voices dominate the world of voiceovers (Bartsch, Burnett, Diller, & Rankin-Williams, 2000), several studies indicate that female voices may be a better choice under some circumstances and that the gender of the voice interacts with the product. Whipple and McManamon (2002) tested listeners' attitudes toward male-gendered, female-gendered, and neutral-gendered products that differed in the voice of the spokesperson. Their results indicated that the gender of the spokesperson

does not have an effect on gender-neutral or male-gendered products. However, for female-gendered products, a female voice elicited a more positive attitude toward the ad. The only scenario where a male voice was preferred was for the female-gendered product when men were the target audience (e.g. for men purchasing the product as a gift). Thus, Whipple and McManamon conclude that female voices have at least the same effectiveness as male voices, if not more.

In examining the gender of spokes-characters (non-human animated characters), Peirce (2001) found that the likelihood that a viewer would buy the product was increased when the gender of the spokes-character matched that of the product (golf balls vs. vacuum cleaners, in the case of this study). Conversely, the gender-neutral spokes-character was not the most effective for the gender-neutral product; instead, the female spokes-character was preferred for the gender-neutral product (coffee). These studies demonstrate that there is little basis to continue to prefer male-gendered voices or spokes-people in advertising.

A third consideration in selecting the voice for an ad is the nativeness of the talker. Although there may be other considerations such as the intended audience or product congruity (e.g., using an Italian-accented voice for pasta), several studies have shown that foreign-accented voices are less intelligible than native voices. In a study examining the effects of voice on listeners' ability to comprehend and retain information from a short narrative, Mayer, Sobko, and Mautone (2003) found that listeners performed better on both a retention task and a transfer task when the speaker was a native speaker of English compared to when the speaker was a second-language learner of English with a Russian accent. They also found that the native speaker received higher positive ratings scores than the nonnative speaker. Foreign accented speech has also been found to be less intelligible when mixed with noise (Lane, 1963; Munro, 1998) and requires more effort to process (Munro & Derwing, 1995).

In the advertising literature, foreign-accented voices have also been shown to elicit less favorable responses and lower purchase intentions. Tsalikis, DeShields, and LaTour (1991) found that Greek-accented English voices received lower scores on 15 bipolar adjectives than native-English voices for a hypothetical commercial for a VCR. In a similar study, DeShields, Kara and Kaynak (1996) tested listeners' attitudes towards native English and Spanish-accented speech when presented with an ad for car insurance. They found that the intent to buy was significantly higher when the speaker was native than for the Spanish-accented speakers. DeShields and de los Santos (2000) found that the impact of accent depends on the relationship between the source of the accent and the listeners. In accordance with previous work, they found that US listeners perceived the native English speaker more positively than the Spanish-accented speaker in an ad for car insurance. Mexican listeners, however, did not rate the native Spanish salesperson differently than the English-accented Spanish-speaking salesperson. DeShields and de los Santos hypothesized that this may be due to the influence the US has on Mexican culture.

Time is Money

Speaking rate is another indexical property that can be manipulated and controlled by advertisers. Because advertising time is expensive, a reasonable question to ask is whether the fast presentation of information, which allows more information to be transmitted in a shorter period of time, has any deleterious effects on listeners' attitudes toward the message, their ability to remember the product, or their intention to buy. Unfortunately, the studies which have examined the effects of speech rate are not conclusive, although the majority suggests that a faster rate is not problematic.

In some cases, faster rates of speech have been shown to be preferred by listeners. Miller, Maruyama, Beaber, and Valone (1976) conducted two experiments in order to test the effects of speech

rate on listeners' attitudes toward the speaker. In the first experiment, groups of listeners heard a passage about the dangers of coffee at two different speaking rates. In addition to varying rate, they also varied the credibility of the speaker by telling listeners that the speaker was either a locksmith or a biochemist. In a second experiment, listeners heard a passage about hydroponically grown vegetables at two speaking rates and at two levels of message complexity. Listeners answered a series of questions designed to determine their attitude toward the speaker. The results showed that listeners judged the speaker of the faster rate to be more knowledgeable, more persuasive, more objective, and also to have greater intelligence. The effects of speech rate were robust; the faster rate elicited more positive responses in all conditions, regardless of the credibility of the speaker or the complexity of the message. One limitation of this study, however, was that the speaker was asked to vary his speech rate, thus it is very likely that other aspects of the voice were altered as well, such as pitch and amplitude.

LaBarbera and MacLachlan (1979), however, avoided these possible confounds by electronically compressing the speech rate. First, they conducted a series of experiments to test listeners' preference for different speech rates. They compressed and expanded the speech rate without altering the pitch of the voice and asked listeners in a paired-comparison task to select which speech sample they preferred. The results of these studies indicated that listeners preferred a faster than normal speaking rate. In a follow-up study, LaBarbera and MacLachlan tested listeners' attitudes and recall of six radio commercials at both normal and fast speech rates. They found that in all cases, the faster commercial was rated as more interesting and elicited higher brand recall after a two-hour delay. Thus, the faster rate was both preferred by listeners and also resulted in higher retention.

MacLachlan (1982) also reported positive effects of faster speech rates. Four radio commercials were used either in their normal or compressed versions, and listeners rated the speaker along four dimensions: friendliness, knowledge, enthusiasm, and energy. The fast commercials were either rated the same as the normal version or more positively. In this study, then, increasing the speech rate had no negative effects on listeners' attitudes about the speaker.

Other studies have shown mixed effects of altering the speech rate. Schlinger, Alwitt, McCarthy, and Green (1983) found that time compression can sometimes interfere with encoding the content in television commercials. Viewers in this study expressed fewer ideas about one of the two commercials in their study when it was presented at the faster rate, but no significant difference was found for the second commercial. As for listeners' attitudes, six of 52 response statements showed the non-compressed version as receiving more positive responses, although the remaining 46 statements showed no difference. Furthermore, the results showed no significant difference in buying intentions for the normal and time-compressed versions of the commercials. Thus, listeners may encode less information and may have fewer positive responses for some response statements, but this does not seem to affect the likelihood that they will actually purchase the product.

More recently, Megehee, Dobie, and Grant (2003) found mixed results for faster rates of speech. They created five versions of a message about the benefits of using a "SmartCard" (an identification card that also functions as a debit card): normal, time-compressed, pause-compressed, time-expanded, and pause-expanded. Thus, three rates (normal, fast, slow) and two methods of rate alteration were studied. Time-adjusted speech changes the overall rate of the message by compressing or expanding all portions equally, but the tempo remains the same. In pause-adjusted speech, the pauses themselves are either shortened or lengthened; thus the actual presentation rate of the words remains the same, but the tempo of the utterance is altered. When comparing the main effect of rate, Megehee et al. found no difference in the attitude toward the product, message, or speaker, though the faster rate did have more affective responses, while the slower rate had more cognitive responses. The authors also found that at faster rates,

time-compression produced more affective responses and a more favorable attitude towards the speaker than did the pause-compressed version.

Chattopadhyay, Dahl, Ritchie, and Shahin (2003) found different results for the method of rate adjustment. They varied both syllable speed (compression of the actual speech forms) and interphrase pause duration and found that reducing the interphrase pause time had little effect on the way listeners processed the message, suggesting that this might be the preferred method of time compression. Increasing syllable speed, on the other hand, did affect the way listeners processed the message, as revealed by measures of attention and recall. They found, however, that increasing syllable rate can increase persuasion, implying that this might be the preferred method of rate compression.

Although a few results from these studies show that increasing the speech rate has some negative consequences (e.g. fewer cognitive responses), the overwhelming conclusion is that faster rates are not problematic and are in some cases preferred. The best method of compression, however, is less obvious. Megehee et al. showed a clear advantage of overall time-compression, whereas Chattopadhyay et al. showed some superiority for pause-compression. Whatever the method of compression, increases in speech rate appear to be well-tolerated by listeners.

When Voice Characteristics Matter Most

The impact of voice characteristics varies depending on how much involvement and interest the listener has with the message. Gelinas-Chebat and Chebat (2001) conducted a study to examine the contribution of voice characteristics on listeners' attitudes toward an ad by varying the level of involvement. Listeners, who were all university students, heard either a low involvement ad which invited them to visit the local bank to acquire an ATM card or a high involvement ad which invited them to visit the local bank to learn about student loans. The assumption was that students would be more interested in learning about student loans since it could directly affect their financial situation. In addition to varying the level of involvement, voice characteristics were varied orthogonally along two dimensions (intensity and intonation) with two levels each, creating four versions of each message. As predicted, the high involvement message increased the acceptance of the arguments of the message. Additionally, changes in voice characteristics did not have an impact on listeners' attitudes in the high-involvement message. However, in the low-involvement message where listeners did not have an a priori interest in the message, the peripheral characteristics of the message (i.e. the changes in voice characteristics) did have an effect on their attitude toward the message. In other words, when listeners do not have a particular interest in the product or message, the quality of the voice that is used has an effect on listeners' attitudes.

Further support for effects of voice on processing and memory comes from another study by Goldinger (1996). He varied the level of processing (LOP) in order to determine whether the focus of listeners' attention would interact with changes in a speaker's voice. In the study phase, listeners encoded 150 words in terms of the gender of the speaker (shallowest LOP), their initial sound, or their syntactic class, namely noun, verb, or adjective (deepest LOP). In the test phase, listeners were given a set of 300 words and were asked to classify words as old or new depending on whether they had been heard in the initial part of the study. Half of the old words were repeated in the same voice and half in a different voice. The strongest effect of voice change was found at the shallowest LOP where words repeated in the same voice received more accurate responses. This result suggests that when listeners' attention is directed toward the deeper symbolic content of the message, they are less disrupted by changes in voice in later recognition tasks. On the other hand, if listeners are not encoding the meaning of the words, but instead are processing them in a shallower manner, inconsistencies in the voice have a significant effect

on their recognition accuracy. Thus, the initial level of encoding of the spoken words determines how much of an impact the voice will have on memory tasks following acquisition.

Integrating Psycholinguistic and Advertising Research

The ultimate goals of advertising are to increase brand recall, instill confidence in the brand, and finally to sell a product. Because much advertising relies on auditory input using spoken language to transfer information about the product to the target audience, the effects of the speech input must be carefully considered. Based on a number of studies in psycholinguistics, speech perception, and marketing research, several general conclusions can be drawn as to how to best control for and manipulate the effects of voice on listeners' attitudes toward and memory for a product.

A natural first question to ask is whether it is important to be selective when choosing a voice for an ad. Two factors related to the encoding of speech make it clear that voice characteristics are crucial for advertising. First, advertising frequently targets a listener's implicit memory for voices since potential consumers are not generally asked to make explicit, direct judgments about the speaker when confronted with a television or radio commercial. Psycholinguistic research demonstrates that voice information that is encoded implicitly lasts at least up to a week in memory. Since advertising tends to target a listener's implicit memory for voices, voice changes and voice characteristics may have both short-term and long-term effects on the success or failure of a marketing campaign.

The second factor which illustrates the importance of voice characteristics in advertising relates to the level of processing. If listeners already have a vested interest in the product, differences in the voice may not affect listeners' perceptions very much. However, when listeners are not personally invested in the content of the message, the vocal characteristics of the talker have significant effects on their attitudes toward the message. Similarly, when listeners encode stimuli in a shallow manner, voice effects are most apparent. Since advertisers are interested in both retaining current consumers and gaining new ones, they cannot guarantee that the listener will have a prior interest in the product. Therefore, voice characteristics are likely to influence the initial encoding of the message and carryover to the buying intentions of potential consumers.

The research reviewed in this chapter establishes a reliable benefit of voice consistency, revealed by a same-voice advantage. Listeners are faster and more accurate when performing linguistic and memory tasks if the voice of the speaker remains constant. Thus, in advertising, it would be advantageous to use a consistent mapping between a voice and a set of ads for a given product. In addition to consistency, familiarity with voices provides a facilatory effect on a range of language processing tasks. Psycholinguistic research reveals that the intelligibility of a talker's voice in noise is better when listeners are familiar with the speaker. This finding is directly relevant for advertising because many commercials are likely to contain music or may be heard in noisy environments (e.g., in a car). Thus, if the listening environment is not ideal and contains conditions that make perceiving the speech more difficult, having a familiar voice can mitigate these factors. Advertising research shows that brand recall is higher when the voice is a celebrity, and therefore familiar to the listener. Finally, the voice of the spokesperson should of course be highly intelligible. Research has shown that nonnative speakers are less intelligible than native speakers and are thus likely to make less ideal candidates for voice advertising, unless other factors, such as product congruity, are relevant.

A practical concern for advertisers is the cost of air time. A realistic concern is whether speech rate can be increased without causing negative effects on listeners' attitudes and memory for the product. In this area, the evidence is promising. Studies of the effects of speech rate on listeners' attitudes and

memory suggest that in general, increasing the rate has no negative effects, and may in fact be preferred. However, considerations of speech rate are not independent of other concerns of voice and linguistic processing. If advertisers elect to use a faster rate of speech, they must be aware of the possible consequences on the behavior of the intended audience. The psycholinguistic research reviewed here shows that at fast presentation rates, consistency of the voice is more important than at slower rates. Thus, if advertisers use a fast rate, they should be sure to use only a small number of voices. Additional research shows that consistency of the voice is less important if listeners are familiar with the speakers. Thus, if advertisers use well-known "celebrity voices" then it may be possible to use more or varied voices in the ad.

Conclusion

Advertisers have a great deal of control over both the linguistic information of an advertising campaign, as well as the indexical information encoded in the speech signal. Evidence from psycholinguistic studies indicates that voice characteristics have an effect on the processing (encoding, storage, retrieval, and transfer) of linguistic information in the message. Thus, it is not only important that advertisers display care when selecting the particular words and content of the message, but also when choosing a voice to represent a specific marketing campaign. If possible, the voice should remain constant across repetitions of an ad, be familiar (either famous or familiar as the result of repetition), and be produced by a native speaker of the language. It is not surprising that these aspects of a speaker's voice affect language processing; these findings are consistent with psychological research on human factors and ergonomics which shows that response consistency, repetition, and familiarity are important for learning and retention. The rate of speech of an ad may be increased without deleterious effects, although in this case, it is even more important that the voice remain consistent. Because consumers of advertising may only be passively attending to a particular ad, selection of the right voice is even more important in these cases where the effects of voice quality have been found to be most apparent.

References

Abercrombie, D. (1967). *Elements of general phonetics.* Chicago: Aldine Publishing Company.

Alsop, R. (1987). Listen closely: These TV ads might have a familiar ring. *The Wall Street Journal, 22* October.

Bartsch, R.A., Burnett, T., Diller, T.R., & Rankin-Williams, E. (2000). Gender representation in television commercials: updating an update. *Sex Roles, 43*, 735-743.

Chattopadhyay, A., Dahl, D.W., Ritchie, R.J.B., & Shahin, K.N. (2003). Hearing voices: The impact of announcer speech characteristics on consumer response to broadcast advertising. *Journal of Consumer Psychology, 13,* 198-204.

DeShields, O.W. Jr., & de los Santos, G. (2000). Salesperson's accent as a globalization issue. *Thunderbird International Business Review, 42*, 29-46.

DeShields, O.W. Jr., Kara, A., & Kaynak, E. (1996). Source effects in purchase decisions: The impact of physical attractiveness and accent of salesperson. *International Journal of Research in Marketing, 13,* 89-101.

Drew, D.G., & Cadwell, R. (1985). Some effects of video editing on perceptions of television news. *Journalism Quarterly, 62*, 828-31, 849.

Gelinas-Chebat, C., & Chebat, J-C. (2001). Effects of two voice characteristics on the attitudes toward advertising messages. *The Journal of Social Psychology, 132,* 447-459.

Glisky, E.L., Polster, M.R., & Routhieaux, B.C. (1995). Double dissociation between item and source memory. *Neuropsychology, 9,* 229-235.

Goldinger, S.D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1166-1183.

Goldinger, S.D., Pisoni, D.B., & Logan, J.S. (1991). On the nature of talker variability effects on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 152-162.

Hilford, A., Glanzer, M., Kim, K., & DeCarlo, L.T. (2002). Regularities of Source Recognition: ROC Analysis. *Journal of Experimental Psychology: General, 131*, 494-510.

Hirahara, T., & Kato, H. (1992). The effect of F0 on vowel identification. In Y. Tohkura, E. Vatikiotis-Bateson & Y. Sagisaka (Eds.), *Speech perception, production and linguistic structure* (pp. 89-112). Tokyo: Ohmsha Publishing.

Kreiman, J., & Van Lancker, D. (1988). Hemispheric specialization for voice recognition: Evidence from dichotic listening. *Brain and Language, 34*, 246-252.

LaBarbera, P., & MacLachlan, J. (1979). Time-compressed speech in radio advertising. *Journal of Marketing, 43*, 30-36.

Landis, T., Buttet, J., Assal, G., & Graves, R. (1982). Dissociation of ear preference in monaural word and voice recognition. *Neuropsychologia, 20*, 501-504.

Lane, H. (1963). Foreign accent and speech distortion. *Journal of the Acoustical Society of America, 35*, 451-453.

Leung, L., & Kee, O.K. (1999). The effects of male celebrity voice-over and gender on product brand name recall, comprehension, and purchase intention. *The New Jersey Journal of Communication, 7*, 81-92.

Lightfoot, N. (1989). Effects of talker familiarity on serial recall of spoken word lists. In *Research on Speech Perception Progress Report No. 15* (pp. 419-443). Bloomington, IN: Speech Research Laboratory, Indiana University

MacLachlan, J. (1982). Listener perception of time-compressed spokespersons. *Journal of Advertising Research, 22*, 47-51.

Mayer, R.E., Sobko, K., & Mautone, P.D. (2003). Social cues in multimedia learning: Role of speaker's voice. *Journal of Educational Psychology, 95*, 419-425.

Megehee, C.M., Dobie, K., & Grant, J. (-2003). Time versus pause manipulation in communications directed to the young adult population: Does it matter? *Journal of Advertising Research, 43*, 281-292.

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

Misra, S., & Beatty, S.E. (1990). Celebrity spokesperson and brand congruence: An assessment of recall and affect. *Journal of Business Research, 21*, 159-173.

Munro, M.J. (1998). The effects of noise on the intelligibility of foreign-accented speech. *Studies in Second Language Acquisition, 20*, 139-154.

Munro, M.J., & Derwing, T.M. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech, 38*, 289-306.

Mullennix, J.W., & Pisoni, D.B. (1990) Stimulus variability and processing dependencies in speech perception. *Perception and Psychophysics, 47*, 379-390.

Nakamura, K., Kawashima, R., Sugiura, M., Kato, T., Nakamura, A., Hatano, K., Nagumo, S., Kubota, K., Fukuda, H., Ito, K., & Kojima, S. (2001). Neural substrates for recognition of familiar voices: a PET study. *Neuropsychologia, 39*, 1047-1054.

Nygaard, L.C., & Pisoni, D.B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics, 60,* 355-376.

Nygaard, L.C., Sommers, M.S., & Pisoni, D.B. (1994). Speech perception as a talker-contingent process. *Psychological Science, 5*, 42-46.

Palmeri, T.J., Goldinger, S.D., & Pisoni, D.B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 309-328.

Peirce, K. (2001). What if the Energizer *Bunny* were female?: Importance of gender in perceptions of advertising spokes-character effectiveness. *Sex Roles, 45*, 845-858.

Plapler, L. (1974). Some famous spokesmen—and how to use them in ads. *Advertising Age*, April 15, 38.

Schacter, D.L. & Church, B.A. (1992). Auditory priming: Implicit and explicit memory for words and voices. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 915-930.

Schlinger, M.J.R., Alwitt, L.F., McCarthy, K.E., & Green, L. (1983). Effects of time compression on attitudes and information processing. *Journal of Marketing, 47*, 79-85.

Shah, N.J., Marshall, J.C., Zafiris, O., Schwab, A., Zilles, K., Markowitsch, H.J., & Fink, G.R. (2001). The neural correlates of person familiarity: A functional magnetic resonance imaging study with clinical implications. *Brain, 124*, 804-815.

Stevens, A.A. (2004). Dissociating the cortical basis of memory for voices, words and tones. *Cognitive Brain Research, 18*, 162-171.

Tsalikis, J., DeShields, O.W. Jr., & LaTour, M.S. (1991). The role of accent on the credibility and effectiveness of the salesperson. *Journal of Personal Selling & Sales Management, 11*, 31-41.

Van Lancker, D.R., Cummings, J.L., Kreiman, J., & Dobkin, B.H. (1988). Phonagnosia: A dissociation between familiar and unfamiliar voices. *Cortex, 24*, 195-209.

Van Lancker, D., & Kreiman, J. (1987). Voice discrimination and recognition are separate abilities. *Neuropsychologia, 25*, 829-834.

Van Lancker, D.R., Kreiman, J., & Cummings, J. (1989). Voice perception deficits: Neuroanatomical correlates of phonagnosia. *Journal of Clinical and Experimental Neuropsychology, 11*, 665-674.

Whipple, T.W., & McManamon, M.K. (2002). Implications of using male and female voices in commercials: An exploratory study. *Journal of Advertising, 31*, 79-91.