

Division of Research
Graduate School of Business Administration
The University of Michigan

July 1980

ON USING VOICE ANALYSIS IN MARKETING RESEARCH

Working Paper No. 223

Nancy J. Nighswonger

Claude R. Martin, Jr.

The University of Michigan

FOR DISCUSSION PURPOSES ONLY

None of this material is to be quoted or
reproduced without the express permission
of the Division of Research.

On Using Voice Analysis in Marketing Research

Abstract

Voice analysis has recently received considerable attention in both the marketing/advertising literature and among marketing research practitioners. This article summarizes the concept behind the technique and identifies major concerns with its use. The authors draw upon research from other disciplines, particularly psycholinguistics, psychosomatic medicine, polygraphy and acoustics for their discussion.

On Using Voice Analysis in Marketing Research

Attitudes are constructs which exist in the minds of individuals. Marketers have long been concerned with their ability to document attitudes, and they have developed numerous attitude scaling techniques in an attempt to operationally define the measurement of consumer attitudes. The most commonly used tool of attitude measurement is the self-report method (Kinnear and Taylor 1979), where respondents report their beliefs or feelings (either directly or indirectly) and their answers are applied to some pre-determined attitude scale. There have been, however, two major problems with this type of measurement. First, the self-report method relies on the subject's ability and/or willingness to provide accurate information. Second, while the self-report method may tap the affective component of a subject's response, the degree of emotional commitment behind the reported attitude can only be measured by using some sort of ordinal scaling, such as a Likert scale or a semantic differential scale. This type of measurement is lacking in objectivity, since the researcher has no way of insuring that each point on the scale represents the same thing for each respondent.

Physiological measurements of beliefs and feelings, on the other hand, operate on the premise that we express with our bodies that which we feel in our minds. Emotional changes resulting from external or internal stimuli are believed to cause various physiological reactions, such as changes in blood pressure, heart rate, pupillary dialation or rate of perspiration. Since these bodily reactions are involuntary, there is no chance for the subject to deliberately distort the response. Although physiological measurements can objectively register a subject's emotional responses to various stimuli, a major

limitation to the physiological response approach in marketing applications has been that it measures only the intensity of feelings and not the direction (positive or negative). Another difficulty with most instruments for measuring or recording physiological functions is that they can only be used while the subject sits quietly in an office or laboratory. Blood pressure, respiration and pulse rate, the psychogalvanic reflex, etc. cannot be recorded while a person is engaged in the normal activities of life. The laboratory situations, besides being highly artificial, are also quite time consuming and expensive. These limitations have historically inhibited the use of most physiological measurements in marketing research, despite their ability to objectively record the intensity of a subject's emotional responses.

Because it contends to overcome these obstacles, voice pitch analysis has received much recent attention by marketing researchers (Brickman 1976, 1980; Haas 1979; Nelson and Schwartz 1979). Voice pitch analysis examines changes in the relative vibration frequency of the human voice in order to measure emotion. The measurement of the pitch changes requires only a tape recorded vocal response, so no complicated or cumbersome equipment is needed at the interview site. Also, since the recording of the physical phenomenon (voice pitch) occurs simultaneously with the subject's conscious interpretation of the attitude (verbal response), the direction (positive or negative) of the attitude can be ascertained from the subject's self-report, while the intensity of the emotion is measured at the same time by mechanical means. In light of the current popularity of voice pitch analysis, the purpose of this paper is two-fold: (1) to discuss some of the methodological problems involved in applying this technique to marketing situations, and (2) to examine the potential of voice pitch analysis in marketing research.

Applications of the Technique

The analysis of voice patterns has been used in many different disciplines for many different purposes. Perhaps the most well-known application of voice analysis is its role in lie detection (see, for example, Kubis 1974), but it has been widely used in other fields as well. Voice analysis has been used, for example, in Psychosomatic research to measure the influence of noise on emotional states (Mason 1969). It has been used in Psycholinguistics to examine the production of sound (Cairns and Cairns 1975; Lieberman 1967; Liberman and Michaels 1962; Ostwald 1963), and in Speech and Music to monitor voice and articulation (Fairbanks 1960; Bradley 1916). Russian researchers have attempted to use voice analysis in order to keep tabs on the stress levels of cosmonauts during space flights (Simonov and Frolov 1973), and voice analysis techniques have been used in studies dealing with stage fright, stress among dental patients, and diagnosis of alcoholics (see Holden 1975). While the uses of voice analysis are diverse, the basic concept is the same: a baseline pitch level is determined for an individual, and deviations from the baseline are examined and interpreted in different contexts.

The Concept

A person's body normally functions at a certain physiological pace, be it fast, slow, or somewhere in-between. For each individual, this normal pace is recorded and referred to as the "baseline." In voice analysis, the normal, or baseline pitch of an individual's speaking voice is charted by engaging the respondent in unemotive conversation. A deviation from the baseline level indicates that the respondent has reacted to a verbal stimulus, such as a question (Ferguson 1973, p. 145).

These vocal "reactions" are explained biologically through the complex role played by the sympathetic division of the autonomic nervous system on different bodily functions. In other words, as long as the respondent can hear and understand the question, the meaning of what the subject has heard and understood will be involuntarily transmitted to a brain center and evaluated. The respondent's brain center sends out a series of nerve impulses after processing the meaning of the question, causing physiological reactions to the emotional stimulus if one has, in fact, occurred. One of the possible reactions is a change in the tension on the vocal cords, causing a change in the pitch level of the voice. These changes in pitch may not be clearly discernable to the human ear, but by plotting the voice inflections in subsequent computer analysis of the recorded voice, even subtle changes in pitch can be detected.

Briefly, then, it is recognized that adults tend to modulate their voices, unconsciously, in sympathy with the emotional meaning of words (Mason 1969, p. 278). Voice pitch analysis attempts to examine the speech signals that reflect these unconscious voice patterns, since vocal cord activity is likely to be influenced by physiological changes resulting from the emotional state of the speaker (Williams and Stevens 1972, p. 1239).

A Caution

The voice pitch methodology is not as straightforward as it might appear on the surface, however. There are many problems which have to be worked out with regard to the manner in which the voice is recorded, the topics or questions to be used in establishing the baseline level for each respondent, and the subsequent interpretation of the noted reactions. For example, there may

be interpersonal differences in response patterns; that is, the emotional "style" of the individual may bias the voice analysis (Hensel 1971; Atkinson 1976). Past or present health conditions, such as asthma, may intensify the respondent's physiological reactions to emotional stimuli (Hahn 1966; Mason 1969), due to the possible malfunctioning of autonomic feedback mechanisms as a result of the health condition. Finally, although research is being conducted in this area, the exact physiological relationships that cause the changes in voice fundamental frequency have not been clearly determined (Atkinson 1978). Consequently, the ability of voice analysis to measure what it purports to measure is still under scrutiny.

Voice Analysis in Marketing Research

The published research on voice analysis in marketing has demonstrated the use of the technique in a number of applications. Brickman (1976, 1980) used the technique three ways: (1) in package research, (2) to predict consumer brand preference for dog food, and (3) to determine which consumers from a target group would be most predisposed toward trying a new product. Nelson and Schwartz (1979) applied voice analysis to measure consumers' emotional responses to advertising, and they compared the predictive power of the voice response scale to the 10-point attitude scale. In all of these studies, the predictions made from voice analysis have been found to reflect actual results better than predictions made using more conventional methods of questioning. Making such predictions, however, without being able to explain the reason behind their increased accuracy is not theoretically sound. The observed relationships may very well be spurious, in which case the technique would be masking rather than uncovering the variable(s) critical to the research being undertaken.

The marketing research, advertising, and other firms that employ voice analysis in marketing contexts have neglected to attend to some of the fundamental issues surrounding the application of the technique. Specifically, there are four major problems left unaddressed in the marketing literature. These areas concern (1) the practice of filtering "lip service" responses out of the sample, (2) the dichotomous nature of the voice pitch scale, (3) the lack of a zero point in the validation studies, and (4) the method of questioning employed in the research. Each of these issues will be discussed individually in the following section. Unfortunately, the studies offered in the marketing literature also contain some measurement and conceptual problems not solely related to the voice analysis technique. For example, in at least one study there is some indication that experimental sensitization may have seriously biased the results. While research problems such as this one should not be overlooked, they are secondary to the main purpose of this discussion: to examine the fundamental issues surrounding the application of the voice analysis technique in marketing research.

Methodological Concerns

"Lip Service" Responses

One of the advantages of voice analysis in marketing is the ability of the technique to filter out "lip service" responses from consumers. A lip service response is defined as a response given by the subject that is something less than a committed response; the subject could be giving the "socially acceptable" response, a confused response, or a response that will merely suffice to fulfill the obligation to the question. Voice analysis procedures filter out these so-called lip service responses, and only deal with those subjects who show an emotional commitment (a significant voice pitch change) to the attitude

reported to the interviewer. According to the VOPAN Marketing Research firm of Boston, Mass. (VOPAN 1980, p. 4), in the majority of cases outside stimulations or internal thoughts do not cause emotions, and subjects in these cases will express attitudes that fall into the category of lip service responses. It is also important to note that, in instances where the respondent is unduly confused or lying (either consciously or unconsciously), the recorded pitch change in their responses will be extremely large. These responses are also filtered out of the system, although Nelson and Schwartz (1979) reported that no more than 5% of their samples to date had shown such extreme voice changes. Eliminated, too, are the subjects whose widely fluctuating vocal characteristics indicate that they feel uncomfortable in the interview situation (Brickman 1976, p. 45).

It becomes apparent, then, that the samples ultimately used in these voice analysis studies are heavily pre-screened. For example, in the study that dealt with predicting consumer brand preferences for dog food (Brickman, 1976), the reported $n = 100$ does not indicate 100 total subjects. Rather, it indicates 100 people who habitually used one of six brands of dog food, felt comfortable being interviewed, had an emotional (voice pitch) reaction to at least six out of 15 pre-determined dog food attributes, successfully completed two interviews (that were scheduled two days apart), and returned a usable questionnaire. The number of subjects filtered out of this study was not reported.

There are three questions that arise as a result of this filtering process. The first question is clearly one of representativeness of the sample; i.e., why do some subjects register emotion on the voice pitch scale while others do not? Are those who are included in the sample somehow different from those who are filtered out of the sample?

The second question that arises deals with the timing of the emotional measurement. When the questions are asked immediately after the stimulus has been presented, the likelihood that the thoughts reported will be related to the stimulus material is no doubt increased (Calder and Sternthal 1980, p. 185). However, if voice analysis is being used to test the effectiveness of advertising messages, for example, filtering out the majority of the subjects simply because they did not register an emotional commitment to purchasing the product (visiting the dealer, etc.) after one exposure to the advertisement seems somewhat harsh. All advertising is not designed to produce an immediate action response on the part of those exposed to it (see Kotler 1980). Prior familiarity with the product or product category, the perception of an immediate need for the product, of differences among respondent's decision making habits could all be factors in determining which subjects register voice pitch changes immediately after having been exposed to one commercial message.

Third, in this filtering process, subjects who are confused, lying, uncomfortable, or truly uncommitted in their responses are eliminated from further analysis. Given, then, that the kinds of responses the voice analysis technique can deal with are limited, it would seem to follow that the kinds of topics that can be used with the voice analysis technique are also limited. A topic that is risky or complicated may provoke a confused response; a sensitive topic may cause the subject to feel uncomfortable and/or to lie with a socially acceptable answer; a highly emotional topic may cause extreme emotional involvement and, accordingly, extreme voice pitch change; and an unemotional topic may induce no significant voice pitch change from the respondent. While it appears to be clearly defined what the voice analysis technique cannot measure, it is vague--at least in a marketing sense--what the technique actually does gauge.

The Voice Pitch Scale

Each individual has a normal, or baseline speaking voice as well as a range within which his voice can fluctuate. Remember that an emotional response is reflected physiologically by a change in the level of voice pitch, and that the direction (positive or negative) of that response is determined by the respondent's subjective evaluation of the emotion. The first scale of measurement used in voice analysis, then, deals with relative pitch changes, not absolute pitch levels. When interpreting the voice pitch measurements, however, the researcher not only looks at the degree of change in pitch level (intensity of emotion), but the direction of the response as well (positive or negative emotion). Hence, the ultimate scale employed shows intensity as well as direction of the recorded emotion; it is a continuous scale that ranges from intense reactions in a negative direction (a "no" response and an extreme voice pitch change) to intense reactions in a positive direction (a "yes" response and an extreme voice pitch change).

To date, marketing researchers have only dealt with the positive end of this voice pitch scale, that is, those subjects who respond "yes" to a trial interest question and concurrently register a heightening of voice pitch within a specified range. Those subjects who register an emotionally charged "no" response, an extreme response at either end of the scale, or an insignificant (unemotional) response are eliminated from further analysis. This practice, though, does not test the capabilities of voice analysis. Rather than ignore the subjects who did not score on the positive end of the scale, the researchers should explore their attitudes using some other form of measurement. For example, how many of the subjects who are eliminated from, say, a study in new product research actually go on to purchase the product? In other words,

it seems appropriate, in addition to determining how many purchases voice analysis does predict, to determine how many purchases the technique does not predict. The entire range of possible responses needs to be evaluated and tested for its actual correspondence with consumer behavior before the validity of the technique can be established.

The Zero Point Problem

In the validation studies that have been done, there have been no controls against which to evaluate the performance of the voice analysis measures. The effects of this lack of a "zero point" will become clear through two examples.

First, in the dog food study previously mentioned, all the subjects were identified as using one of six brands of dog food on a habitual basis, but they were instructed not to reveal which of the six brands they actually used. The voice analysis technique was then employed to predict each consumer's actual brand preference. There were, however, no "control" subjects who used none of the six brands on a habitual basis. Consequently (and possibly aggravated by the fact that the subjects were aware of the purpose of the study), the attributes that were determined in the first interview to constitute the subject's "ideal brand" may, in fact, have come subjectively very close to describing the brand the customer actually used. If the subject was already a habitual user of a particular dog food, the purchase decision process would have been reduced to an almost automatic behavior. The degree of familiarity with six different brands of dog food, then, would appear to be limited. The subject's frame of reference for answering the attribute questions would be restricted to past experience, which, theoretically, reduces his range of knowledge about dog food down to one or two brands, given that the subject has been defined as a

habitual user of one particular brand. It would seem, then, that the correct prediction of the subject's actual brand is almost inevitable. If familiarity with a topic has anything to do with the subject's attitude or commitment toward that topic, then it would seem clear that the results of this study may have been seriously biased by the lack of "control" subjects.

This issue of a lack of a zero point in the validation studies is illustrated even more clearly by the work done in determining which of a number of commercial messages should be chosen for an advertising campaign (VOPAN 1980, p. 19). Three commercials that had been pre-tested in test markets were also tested using voice analysis techniques in order to determine which commercial was most effective in generating sales. While the voice analysis scores matched the test market data in terms of the effectiveness rankings of the three commercials (commercial X had a lower voice analysis score than Y, which had a lower score than Z, where Z was clearly the most effective commercial in the test market), the voice analysis scores said nothing about the share of market that was expected to be generated by the three commercials. There was no zero point; no "choose none of the three" category. It is possible that, out of the three commercials, none were expected to generate enough sales to justify the advertising expenditure. Market-test data may provide an indication of some problem along this line, whereas voice analysis data would not.

Questioning

There is evidence to indicate that once an emotional response has been registered in the voice, it takes some time before the pitch of the voice goes back to the normal baseline level (Friedhoff, Alpert and Kurtzberg 1973; Dunbar 1954). By not including neutral (unemotive) questions between the questions

designed to elicit an emotional reaction, the control "baseline" is not re-established after each emotional response that is registered and, hence, it cannot be used as a neutral comparison ground across all responses. Giving the subjects (as the marketers have apparently done) a battery of questions designed to elicit emotional reactions with no neutral questions in-between makes a static formula for calculating deviations in pitch from the mean level impossible to apply accurately.

One Final Point

Voice analysis is a procedure that depends on physiological changes in the body that occur as a result of emotion. Ordinarily, emotion does change bodily states and quantitative relationships in the body, but there is no evidence to indicate that a particular bodily response is associated with any particular emotion. Thus, any given element in the complex network of bodily reactions may, or may not change with emotional states (Dunbar 1954).

This variability in physiological responses may be a function of personal style (Hensel 1971; Williams and Stevens 1972), and/or of the specific emotion being felt by the respondent (Ferguson 1973, p. 157). Whatever the source of this variability, it is clear that one physiological measure, standing alone, is not sensitive to every emotion, nor is it sensitive to any particular set of emotions for every individual. This evidence implies that voice analysis should be used in conjunction with, rather than in place of other methods of attitude measurement.

IMPLICATIONS FOR FUTURE RESEARCH

In light of the lack of conclusive research dealing with the details of the voice as a physiological correlate of emotion, we hypothesize--for the time

being--that voice analysis measures the direction (positive or negative) of an attitude as well as the emotional commitment behind that attitude. Based on this hypothesis, it is clear that immediate research efforts should focus on establishing the concurrent validity of voice analysis as an attitude measure. Many tests to explore this concurrent validity are potentially possible. However, there does appear to be a logical link between voice pitch analysis and response latency research (MacLachlan, Czepiel and LaBarbera 1979; Aaker, Bagozzi, Carman, and MacLachlan 1980).

Response latency is defined as the amount of time a respondent spends in deliberation before answering a question. MacLachlan, et al. (1979) found that response latency is a robust measure that serves as an indicator of the subject's certainty in a given response. Thus, response latency has been shown to be a valid and sensitive measure of the conviction with which attitudes are held. Since data used in voice analysis are recorded on tape, the response latency of the subjects can be easily and inexpensively obtained along with the voice pitch measurements. Combining voice and response latency analyses presents the promise of simultaneously measuring the same phenomenon and, thus, testing for concurrent validity. This link between the two methods is also a prime candidate for the next research step.

Over and above the test for concurrent validity, the combining of voice pitch analysis with response latency research may answer our previous criticism of use of single physiological measure. In other words, voice analysis when combined with the response latency technique may result in a more sensitive and effective measure of attitudes than voice analysis standing alone.

As we have discussed, although there appear to be some problems associated with the application of voice analysis in marketing, the technique does seem to

have promise. Nevertheless, further research is needed in order to more clearly define the limitations and capabilities of the technique before voice analysis can be employed with confidence in the marketing discipline.

References

- Aaker, David A., Richard P. Bagozzi, James M. Carman, and James M. MacLachlan (1980), "On Using Response Latency to Measure Preference," Journal of Marketing Research, 17 (May), 237-244.
- Atkinson, James E. (1978), "Correlation Analysis of the Physiological Factors Controlling Fundamental Voice Frequency," Journal of the Acoustical Society of America, 63(1) (January), 211-222.
- _____ (1976), "Inter- and Intraspeaker Variability in Fundamental Voice Frequency," Journal of the Acoustical Society of America, 60(2) (August), 440-445.
- Bradley, Cornelius Beach (1916), ...On Plotting the Inflections of the Voice, Berkeley: University of California Press.
- Brickman, Glen A. (1976), "Voice Analysis," Journal of Advertising Research, 16(3) (June), 43-48.
- _____ (1980), "Uses of Voice-pitch Analysis," Journal of Advertising Research, 20(2) (April), 69-73.
- Cairns, Helen and Charles Cairns (1975), Psycholinguistics: A Cognitive View of Language, New York: Holt, Rinehart and Winston.
- Calder, B. J. and B. Sternthal (1980), "Television Commercial Wearout: An Information Processing View," Journal of Marketing Research, 17 (May), 173-186.
- Dunbar, Flanders (1954), Emotions and Bodily Changes, 4th edition, New York: Columbia University Press, 158.
- Fairbanks, Grant (1960), Voice and Articulation Drillbook, 2nd edition, New York: Harper.
- Ferguson, Robert J. and Alan L. Miller (1973), The Polygraph in Court, Springfield, IL: Thomas Books.
- Friedhoff, Arnold J., Murray Alpert and Richard L. Kurtzberg (1973), "An Effect of Emotion on Voice," Nature, 193(4813), 357-358.
- Haas, Charlie (1979), "Charlie Haas on Advertising," New West, (November 5), 31-59.
- Hahn, W. W. (1966), "Autonomic Responses of Asthmatic Children," Psychosomatic Medicine, 28, 323-332.

- Hensel, James Stephen (1971), Physiological Measures of Advertising Effectiveness: A Theoretical and Empirical Investigation, doctoral dissertation, The Ohio State University, Ann Arbor: University Microfilms International, 20.
- Holden, Constance (1975), "Lie Detectors: PSE Gains Audience Despite Critics' Doubts," Science, 190 (October 24), 359-362.
- Kinnear, Thomas C. and James R. Taylor (1979), Marketing Research: An Applied Approach, New York: McGraw-Hill Book Company, 301.
- Kotler, Philip (1980), Principles of Marketing, Englewood Cliffs, NJ: Prentice-Hall, Inc., 502.
- Kubis, Joseph A. (1974), "Comparison of Voice Analysis and Polygraph as Lie Detection Procedures," Polygraph, 3 (March), 1-41.
- Lieberman, Philip (1967), Intonation, Perception and Language, Research Monograph No. 38, Cambridge, Mass.: The M.I.T. Press.
- _____ and S. Michaels (1962), "Some Aspects of Fundamental Frequency, Envelope Amplitude, and the Emotional Content of Speech," Journal of the Acoustical Society of America, 34, 922-927.
- MacLachlan, James, John Czepiel, and Priscilla LaBarbera (1979), "Implementation of Response Latency Measures," Journal of Marketing Research, 16 (November), 573-577.
- Mason, R. K. (1969), "The Influence of Noise on Emotional States," Journal of Psychosomatic Research, 13, 275-282.
- Nelson, Ronald G. and David Schwartz (1979), "Voice-Pitch Analysis," Journal of Advertising Research, 19(5) (October), 55-59.
- Ostwald, Peter F. (1963), Soundmaking: The Acoustic Communication of Emotion, American Lecture Series No. 538, Springfield, IL: Thomas Books.
- Simonov, P. V. and M. V. Frolov (1973), "Utilization of Human Voice for Estimation of Man's Emotional Stress and State of Attention," Aerospace Medicine, 44, 256-258.
- VOPAN Marketing Research (1980), "VOPAN[®] Voice Pitch Analysis in Marketing and Advertising Research," unpublished company literature, The VOPAN Company, Boston, Mass., (January).
- Williams, Carl E. and Kenneth N. Stevens (1972), "Emotions and Speech: Some Acoustical Correlates," Journal of the Acoustical Society of America, 52(4), 1238-1250.