DOCUMENT RESUME

ED 028 442 48 AL 001 844

By-Vanderslice, Ralph
The "Voiceprint" Myth.
Michigan Univ., Ann Arbor. Center for Research on Language and Language Behavior.
Spons Agency-Office of Education (DHEW), Washington, D.C. Bureau of Research.
Bureau No-BR-6-1784
Pub Date 1 Feb 69
Contract-OEC-3-6-061784-0508
Note-21p.; Report included in Studies in Language and Language Behavior, Progress Report No. VIII.
EDRS Price MF-$0.25 HC-$1.15
Descriptors-*Acoustic Phonetics, *Articulation (Speech), Auditory Discrimination, Auditory Perception, *Phonetic Analysis, *Spectrograms
Identifiers-*Voiceprint Identification

The technique of "voiceprint identification" has been invested with a myth of infallibility, largely by means of a specious analogy with fingerprints. The refusal of its chief proponent to submit to a properly controlled test of his ability, coupled with the inability o observers in independent studies to get comparably low error rates, is sufficient basis for skepticism. There is reason furthermore to doubt whether spectrograms could ever form a valid basis for absolute identification. The present study replicated "The 'Voiceprint' Game" with even more stringent constraints on selection and arrangement of spectrograms representing three undisguised tokens of the same word or phrase by each of three adult male native speakers of North Midland American English. The results show that (1) most of the similarity between any two spectrograms of the same word is "linguistic," arising from that particular word's being spoken in the same or similar dialect and context; and (2) differences between the spectrograms of the same word spoken by two different speakers of similar dialects are "small" in comparison to within-speaker variations for that word. The "Voiceprinter" is more properly likened to the "lie detector" in that stable guidelines for interpreting its graphic records appear permanently beyond reach. (Author/AMM)

THE "VOICEPRINT" MYTH[1]

Ralph Vanderslice

Center for Research on Language and Language Behavior
The University of Michigan

The so-called technique of "voiceprint identification" has been invested with a myth of infallibitity, largely by means of a specious analogy with fingerprints. The refusal of its chief proponent to submit to a properly controlled test of his ability, coupled with the inability of Os in independent studies to get comparably low error rates, is sufficient basis for skepticism. There is reason furthermore to doubt whether spectrograms could ever form a valid basis for absolute identification. The present study replicated "The 'Voiceprint' Game" with even more stringent constraints on selection and arrangement of spectrograms representing three undisguised tokens of the same word or phrase by each of three adult male native speakers of North Midland American English. The results show that: (a) most of the similarity between any two spectrograms of the same word is linguistic, arising from that particular word's being spoken in the same or similar dialect and context, irrespective of whether one or two speakers were involved; and (b) differences between the spectrograms of the same word spoken by two different speakers of similar dialect are small in comparison to within-speaker variations for that word, even if the speakers are easy to distinguish by ear. When sophisticated vocal disguises are included, intra-S variance for the same word becomes enormous in comparison to normal inter-S differences. The "Voiceprinter" (spectrograph) is more properly likened to the "lie detector" (polygraph) in that stable guidelines for interpreting its graphic records appear permanently beyond reach.

A myth of infallibility has grown up around the so-called technique of "voiceprint identification" as a result of widespread publicity given to the sanguine but unsubstantiated claims of its proponents, especially L. G. Kersta.[2] A method of identifying people from recordings of their voices would be a most useful tool in law enforcement, counterespionage, and many other applications. However, the hypothesis that because each individual's vocal cavities are different, his acoustic output must likewise be unique, is entirely unproven. The artfully nurtured analogy with fingerprints[3] is grossly misleading: intra-S variance of palmar ridge conformations is practically nil from birth to death; large intra-S variance of vocal tract shape and acoustic output is a sine qua non of oral communication.

Attempts to replicate Mr. Kersta's earlier quasi-formal and inadequately reported[4] experiments (Stevens, Williams, Carbonell, & Woods, 1968; Young & Campbell, 1967) have obtained very much higher error rates,[5] but do not thereby directly refute his claim to a personal record of 100% correct identifications.[6] As with phrenology a century earlier, "voiceprint" advocates dismiss any counter-evidence as irrelevent on grounds the dissenting experimenters are non-initiates. Kersta's ostensible infallibility can be directly tested only if he agrees to demonstrate it by taking a test prepared by an independent phonetics laboratory, in which sophisticated vocal disguises[7] and competent mimicry,[8] as well as voices chosen to be naturally similar, are employed. Kersta's resolute refusal to accept such a challenge,[9] coupled with the inability of Os in properly controlled studies to get comparably low error rates, is reason enough to be skeptical of his claims.

Meanwhile one relevant direction for independent research is to assess the likelihood that voice spectrograms could ever form a valid basis for absolute identification, by examining the fundamental issue of intra-S similarity in relation to inter-S differences. "The 'Voiceprint' Game" (Vanderslice, 1966) showed that, among only three Ss, whose voices sounded nothing alike, and who made no attempt at vocal disguise:

(1) A given speaker, saying the same word on different occasions, often produced widely different spectrographic patterns; and

(2) Different speakers often produced spectrographic patterns which were substantially identical.

It goes without saying that no comparable demonstration has ever been arranged with handwriting or fingerprints--or other sorts of admissible evidence (relative to their probative functions), e.g. bullets, tire tracks, blood tests, or typewriting.[10]

Nevertheless, "The 'Voiceprint' Game" has been impugned (in the brief of plaintiff-petitioner on appeal in State of New Jersey v. Paul Grodon Cary) on the grounds that (a) it was "deliberately devised to confuse [i.e. convince]" the observer [p. 62]" (true--this of course was its raison d'être); (b) "the spectrograms used were carefully selected for this [persuasive] purpose [p. 62]" (true but irrelevant since Kersta's "technique" is open to precisely the same charge); and (c) "the clarity of the spectrograms...was somewhat blurred by a copying process [p. 63]" (true--multilithed copies of the spectrograms had to

be cut from a copy of the "Game" for use in the New Jersey case because the originals had been impounded as a defense exhibit in The People v. Edward Lee King, Jr. in Los Angeles.

New Evidence

The following experiment was carried out in order to overcome (without granting the relevance of) these objections, and to see whether the results obtained in the "Game" would be replicable under even more stringent constraints.

Procedure. As before, three adult male Ss who were native speakers of North Midland American English were chosen without further regard for vocal similarity. As before, their voices were recorded on an Ampex tape recorder with a high quality microphone located in a sound-treated room, and the broadband spectrograms were made on a Kay Electric Sona-Graph (this time a model 6061-B) following the same procedures and with approximately the same frequency range. In order to overcome the objection to what freedom there was in preparation of "The 'Voiceprint' Game" to select and arrange the spectrograms in a way favoring the points to be made (intra-S variance is large; inter-S similarity is common), this time only one paradigm set of three tokens of one 2-sec. test utterance was recorded from each S. One S later returned and recorded several additional tokens, with and without attempts at vocal diguise, but these were never used in lieu of his initial set of three.

The aim of these paradigm sets was to illustrate the effects of trivial changes of bodily posture in increasing the inevitable intra-S variance--even where all three tokens were recorded within moments, eliminating the substantial variations occurring at diurnal or slower rates. The bodily postures were as follows: each S spoke the first token sitting upright with head erect, the second leaning forward with head lifted, and the third sitting sideways with head turned toward the microphone. It should be noted that the auditory effect of these changes is practically undetectable.

Results. Figures 1 through 4 present snipped-up spectrogram segements showing single words or short phrases--exactly as are used in the practice of forensic sound spectrography--arranged in matrix format. Opposite each figure in an identifying key with the same tabular layout. In every case the leftmost three columns contain S's paradigm sets, one column per speaker; and, within those columns, the rows are by head position: top row, erect; middle row, raised; and lower row,

turned. The fourth column from the left always contains other undiguised tokens from speaker L (whose paradigm set is in column three) and the fifth and sixth columns, when present, show disguised tokens from the same S. Thus for the first three columns of each figure there were zero degrees of freedom in the selection and arrangement of the spectrograms. Once again it is evident that:

1. Most of the similarity betwen any two spectrograms of the same word is linguistic, arising from that particular word's being spoken in the same or a similar dialect and context, irrespective of whether one or two speakers were involved.
2. Differences between the spectrograms of the same word spoken by two different speakers of similar dialect are small in comparison to within-speaker variations for that word, even if the speakers are easy to distinguish by ear.

Insert Figures 1 through 4 about here

It can further be observed (from the rightmost column in Figure 1 and the rightmost two in Figure 2) that while the vocal disguises employed here varied in their effectiveness, some of them were very effective indeed. A description of these disguise modes is beyond the scope of this paper and must await research in progress on "some parameters of vocal disguise." In general, distortions of tract and timing features are more efficacious than of source features (falsetto, whisper, ingressive voice), as could be predicted from a knowledge of the spectrographic process.

Conclusion

Already, several persons have been sentenced to prison with "voiceprints" as the primary evidence of their guilt. Some may well have been innocent. There is no question but that any bona fide advances in the technology of acoustic phonetics should be made available, as applicable, to the legitimate purposes of the police.[11] However, "voiceprint identification" is to be likened not to fingerprints but rather to polygraph or "lie detector" tests (cf. Smith, 1967). Neither method is accepted by scientists in the relevant disciplines as reliable even in the technical sense of the word--far less its popular sense of "valid."

Both employ large, complicated electronic devices likely to awe the lay juryman--or, still less legitimately, to intimidate suspects into confessing out of fear of a machine which in fact (so far an enabling the identification of people by their voices or the detection of untruths, respectively) is a hoax. The "Voice-printer" (spectrograph) and the "lie detector" (polygraph) are precisely alike in that stable guidelines or standards for interpreting their graphic records, to negotiate the respective binary cruces (same or different voices; true or false responses) and arrive at valid (correct) and reliable (repeatable) decisions, appear permanently out of reach. Admission of such evidence into courts of law as having probative value makes a mockery of that revered jurisprudential principle, presumption of innocence.

References

Cohen, J. The Watts voiceprint case: A TV boast that trapped an arsonist. West [magazine of the Los Angeles Times], March 26, 1967, 12-19.

Ladefoged, P., & Vanderslice, R. The "voiceprint" mystique. Working Papers in Phonetics, 1967, 7, 126-142.

Prince, W. N. Voiceprint: Tool or toy? West [magazine of the Los Angeles Times], March 26, 1967, 14.

Smith, B. M. The polygraph. Scientific American, 1967, 216(1), 25-31.

Stevens, K. N., Williams, C. E., Jr., Carbonell, J. R., & Woods, B. Speaker authentication and identification: A comparison of spectrographic and auditory presentations of speech material. Journal of the Acoustical Society of America, 1968, 44, 1596-1607.

Vanderslice, R. The "voiceprint" game. Working Papers in Phonetics (unnumbered supplement), 1966, 1-12.

Young, M. A., & Campbell, R. A. Effects of context on talker identification. Journal of the Acoustical Society of America, 1967, 42, 1250-1254.

Footnotes

[1]This paper is based in part on a talk entitled "The 'Voiceprint' Hoax" given at the Graduate Center of the City University of New York in December, 1968, and at the State University of New York at Buffalo in February, 1969. The research reported herein was performed in part pursuant to Contract OEC-3-6-061784-0508 with the U. S. Department of Health, Education, and Welfare, Office of Education, under provisions of P. L. 83-531, Cooperative Research, and the provisions of Title VI, P. L. 85-864, as amended. This research report is one of several which have been submitted to the Office of Education as Studies in Language and Language Behavior, Progress Report No. VIII, February 1, 1969.

[2]Mr. Kersta is president and founder of Voiceprint Laboratories, Inc. (now a division of Farrington Manufacturing Co.) which he established upon leaving the Bell Telephone Laboratories (c. 1965) after being a member of the technical staff for 39 years. Exaggerated claims for "voiceprint" accuracy were publicized as early as 1962 under the aegis of Bell Labs in a campaign calculated to discourage obscene and other nuisance phone calls. A full page paid advertisement in (inter alia) Physics Today, December 1962, titled News from Bell Telephone Laboratories, read: "WE'RE 'FINGERPRINTING' VOICES .... Voiceprints... are actual pictures of sound, revealing the patterns of voice energy. Each pattern is distinctive and identifiable. They are so distinctive that voiceprints may have a place, along with fingerprint and handwriting identification, as an important tool of law enforcement.

"The shape and size of a person's mouth, throat and nasal cavities cause his voice energy to be concentrated into bands of frequencies. The pattern of these bands remains essentially the same despite modifications which may result from loss of teeth or tonsils, the advancement of age, or attempts to disguise the voice."

Kersta on The Joe Pyne Show (February, 1967) stated that "we know of no way a person can change his speech such that it is impossible to identify him. It sometimes makes it a little more difficult--it takes longer to do it--but we have not found a condition where it has been impossible to find identifying features in a person's voice." Even more egregious claims were made by L. G. Kersta Jr. on a BBC broadcast (c. 1966): "People cannot disguise their voices, we found this out. We can identify you by your voice. [Each voice is individual]

like your fingerprint.... We've had numerous law enforcement cases: obscene telephone calls, bomb threats; even have your mouth full of marbles, we'll still identify you...or disguise your voice in any way, we'll still pick you up. It doesn't make any difference at all." But in fact all these grandiose claims are based on a few informal pseudo-experiments with the most naive vocal "disguises" imaginable.

[3]The analogy with fingerprints, implicit in the name, is forwarded in various covert ways (besides bald comparisons as cited in Footnote 2). One is the continual use of contour spectrograms in publicity releases, apparently for their exotic look and vague resemblance to fingerprints. Indeed the same six contour prints of the word you (or a subset), in which the two matching ones are of Kersta's voice, have been used ever since 1962 (see Figure A-1, Appendix A). But in fact these displays are not "voiceprints" at all in the sense in which this term is used in forensic sound spectrography, since only wide-band ("bar") spectrograms are used for comparisons by eye.

A second way of promoting the analogy with fingerprints is the incorporation into the "voiceprint" mystique of nomenclature borrowed form the more prestigious art. For example, Kersta makes much of counting up "points of similarity"--although he is unable to define what constitutes one. The closest he has come was drawing sketches alongside each pair of spectrograms introduced by the prosecution in the King trial to aid the uninitiated jurors in seeing the supposed similarities. Figures B-1 and B-2, Appendix B, reproduce these "voiceprints" along with Kersta's sketches. Comparison with between-speakers pairs in Figures 1-4 suggests that his criteria of similarity are rather compliant.

[4]See Stevens, et al., 1968, p. 1606n; also Ladefoged and Vanderslice, 1967, pp. 135-136.

[5]They showed also the clear superiority of ear over eye in this task of speaker recognition, contrary to Kersta's claim.

[6]A record which Kersta bases on his "over 100 field applications" whose accuracy is attested by the fact that, purportedly, charges were always dropped or convictions obtained in accord with his reports. In two crucial cases (King & Cary), convictions based chiefly on "voiceprint" evidence have been reversed for that reason on appeal. But a definitive decision on "voiceprint" admissibility still hanges in the balance, awaiting further litigation.

[7]Not such ridiculous procedures as holding the nose, putting marbles in the mouth, or a handkerchief over the telephone mouthpiece, etc. (cf. Prince, 1967).

[8]Rather than the mere impressions of public figures (e.g., J. F. Kennedy) by entertainers (e.g., Vaughan Meador) which Kersta has pitted his "technique" against (informally).

[9]P. Ladefoged, Professor of Phonetics at UCLA delivered one such challenge when he appeared opposite Kersta on nationwide TV (The Joe Pyne Show, February, 1967). Mr. Kersta's reply to Dr. Ladefoged was: "I have many more important things to do than play kindergarten games with you."

[10]While it is true that even such widely accepted forms of evidence as fingerprints and ballistics require interpretation, yet their probative worth is founded on the legal fiction that a jury can "draw its own conclusion from the evidence," and that the opinion of experts is merely an "aid and guide" to their deliberations. It is difficult to see how this could be maintained in the use of "voiceprints."

It is sometimes suggested that even though "voiceprints" are not trustworthy as evidence to convict, they might be allowed, like blood tests in paternity cases, to exonerate--i.e. prove that two voices could not be from the same person. The data in Figures 1-4, however, show that intra-S variations for the same word or phrase (particularly if a vocal disguise must be allowed for) are so enormous relative to typical inter-S differences that no such conclusion could be warranted. Indeed "voiceprints" admitted for this purpose, albeit less invidious to civil liberties and due process, might (if believed by a jury) undo months of careful police work and result in the specious exoneration of a hardened criminal. Since Kersta claims to have reported negative identifications in "over 80% of the more than 100 field applications [personal communication, November, 1967]," this may well be the gravest danger posed by his mystique.

[11]One of the more unsavory aspects of the King trial in Los Angeles was the patent use of "voiceprints" by the prosecution to circumvent the inadmissibility, under the Miranda decision, of an illegally obtained confession tape. Portions of this tape, including at least one clear instance of an admission, were sent to Kersta as the exemplar of King's voice (see Ladefoged & Vanderslice, 1967, pp. 140-141; also Cohen, 1967).

Appendix A

Figure A-1 shows the hackneyed set of six contour spectrograms of the word you, which have been printed and reprinted ubiquitously. They are not "voice-prints" as that term is applied in law enforcement cases. This set of course conveys a quite misleading impression of the intra-S similarity and inter-S differences to be expected in random sample of population (cf. Figures 1-4).

Appendix B

Figures B-1 and B-2 show the entire body of evidence Kersta presented in the King trial to "prove" that an unknown voice on the audio tape of a CBS pre-TV interview and an unknown voice on the audio portion of the CBS broadcast video-tape, respectively, were both the voice of the defendant. There was no other evidence whatever to show that the two unknown voices were of one and the same person (whether or not it was King). Thus Kersta's claim (in a rebuttal to "The 'Voiceprint' Mystique" at a meeting of the Acoustical Society of America, Miami Beach, November, 1967, and elsewhere) that he has never used fewer than thirteen pairs of words in making an identification is not strictly veridical. The five pairs in Figure B-1 must be considered the minimum (Ladefoged & Vanderslice, 1967) at least until a new low is established.

Figure Captions

Fig. 1. Spectrograms of the phrase be out spoken by Ss J, K, and L. S and other pertinent information, including head position in paradigm sets, noted at corresponding position in key.

Fig. 2. Spectrograms of the word I spoken by Ss J, K, and L. S and other pertinent information, including head position in paradigm sets, noted at corresponding position in key.

Fig. 3. Spectrograms of the word door spoken by Ss J, K, and L. S and other pertinent information, including head position in paradigm sets, noted at corresponding position in key.

Fig. 4. Spectrograms of the phrase know you will spoken by Ss J, K, and L. S and other pertinent information, including head position in paradigm sets, noted at corresponding position in key.

Fig. A-1. Contour spectorgrams of the word you spoken by five Ss. Upper left and lower right represent same S (Kersta). These are not "voiceprints" as that term is used in forensic applications.

Vanderslice

Fig. B-1. People's exhibit 19 from the King case: Kersta's spectrograms, with putative "points of similarity" sketched, purporting to link an unknown voice on the CBS pre-TV interview audio tape (left member of each pair) with the known voice of Edward Lee King, Jr. (right members) illicitly taped during his interrogation. Reproduced unretouched as labeled and mounted by Kersta, except for excision of blank space and slight reduction in size.

Fig. B-2. People's exhibit 20: Kersta's spectrograms and sketches purporting to link an unknown voice on the audio portion of the CBS broadcast videotape (left member of each pair) with the known voice of Edward Lee King, Jr. (right member) illicitly taped during his interrogation. Reproduced unretouched as labeled and mounted by Kersta, except for excision of blank space and slight reduction in size.

J head erect	K head erect	L head erect	L head erect	L head erect
J head raised	K head raised	L head raised	L another normal token	L disguised voice
J head turned	K head turned	L head turned	L another normal token	L disguised voice

Key to Figure 1

Figure 1

J head erect	K head erect	L head erect	L another normal token	L disguised voice	L disguised voice
J head raised	K head raised	L head raised	L another normal token	L disguised voice	L disguised voice
J head turned	K head turned	L head turned	L another normal token	L disguised voice	L disguised voice

Key to Figure 2 . . .

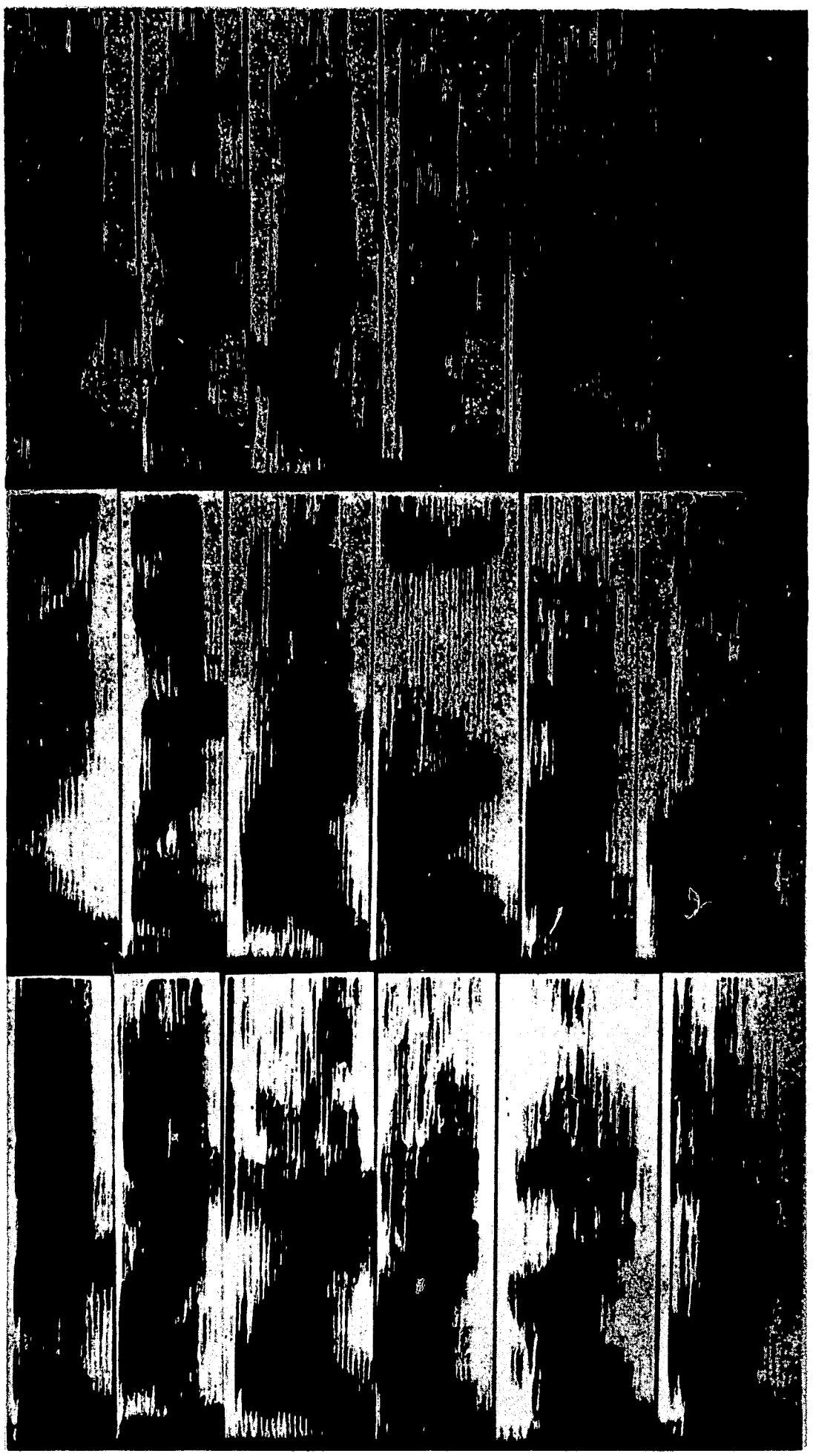

Figure 2

J head erect	K head erect	L head erect	L another normal token
J head raised	K head raised	L head raised	L another normal token
J head turned	K head turned	L head turned	L another normal token

Key to Figure 3

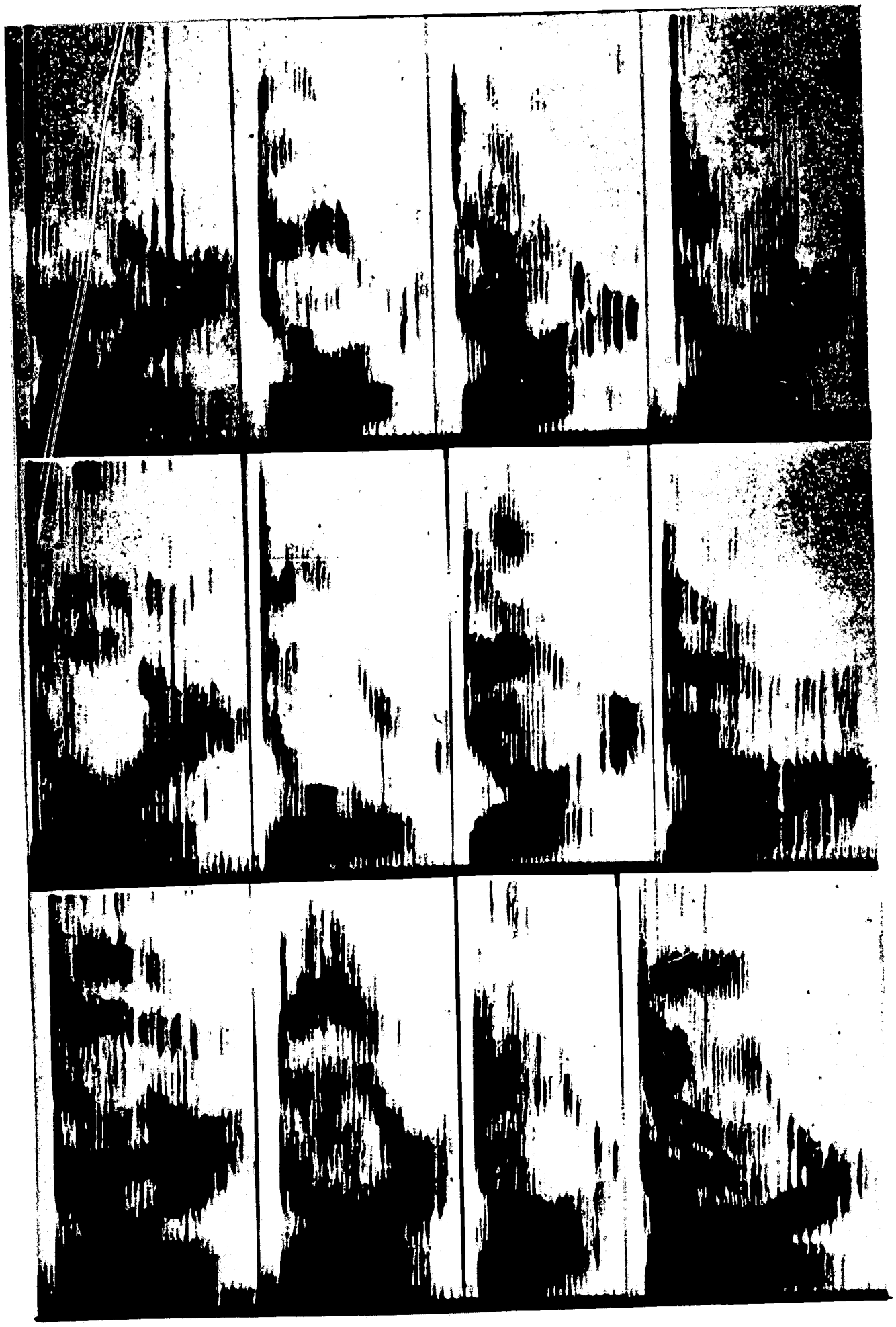

Figure 3

J head erect	K head erect	L head erect	L another normal token
J head raised	K head raised	L head raised	L another normal token
J head turned	K head turned	L head turned	L another normal token

Key to Figure 4

Figure 4

Figure A-1

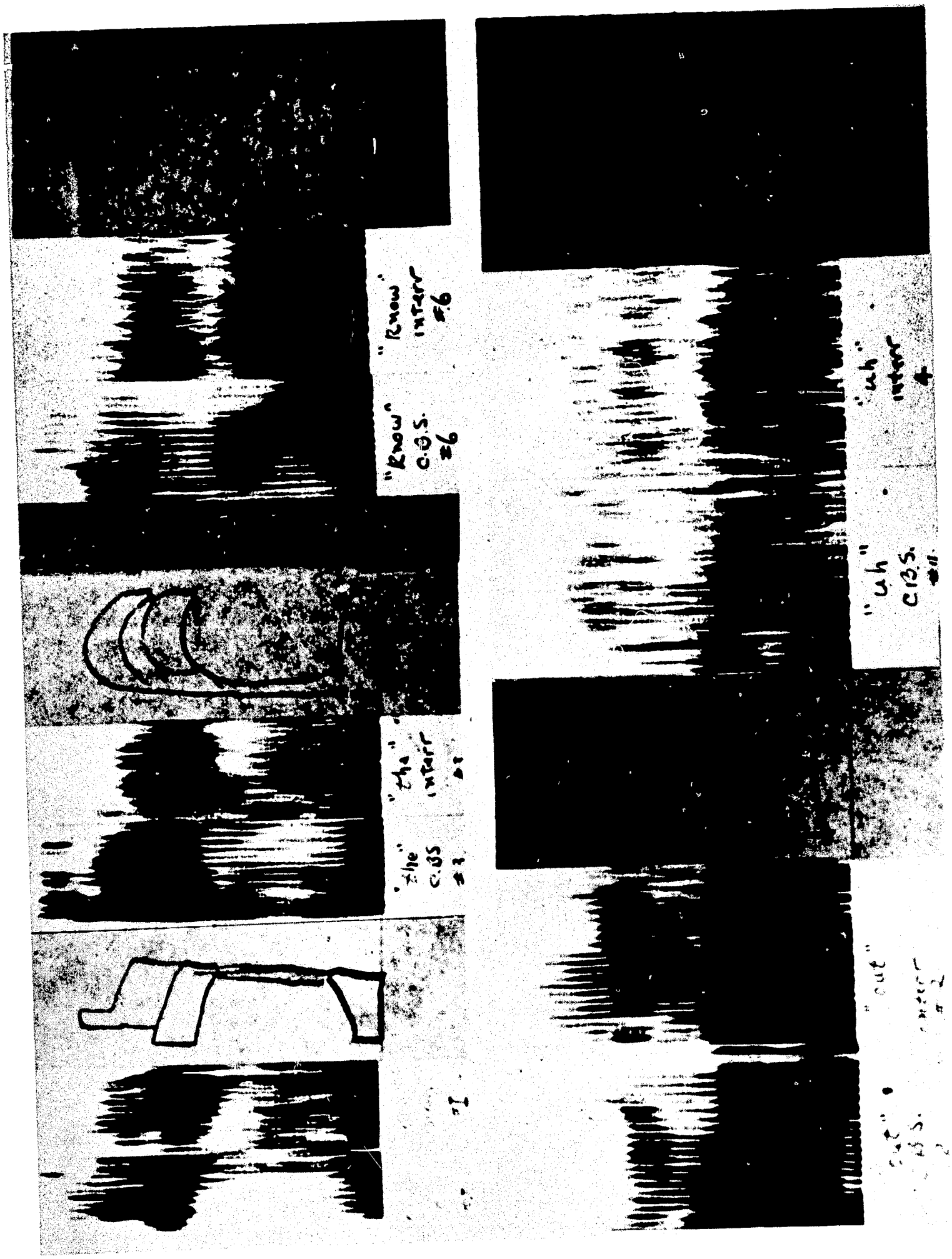
"Know" C.B.S. #6
"Know" Interv #6
"the" C.B.S. #1
"the" Interv #1
"uh" C.B.S. #1
"uh" Interv #1

Figure B-1

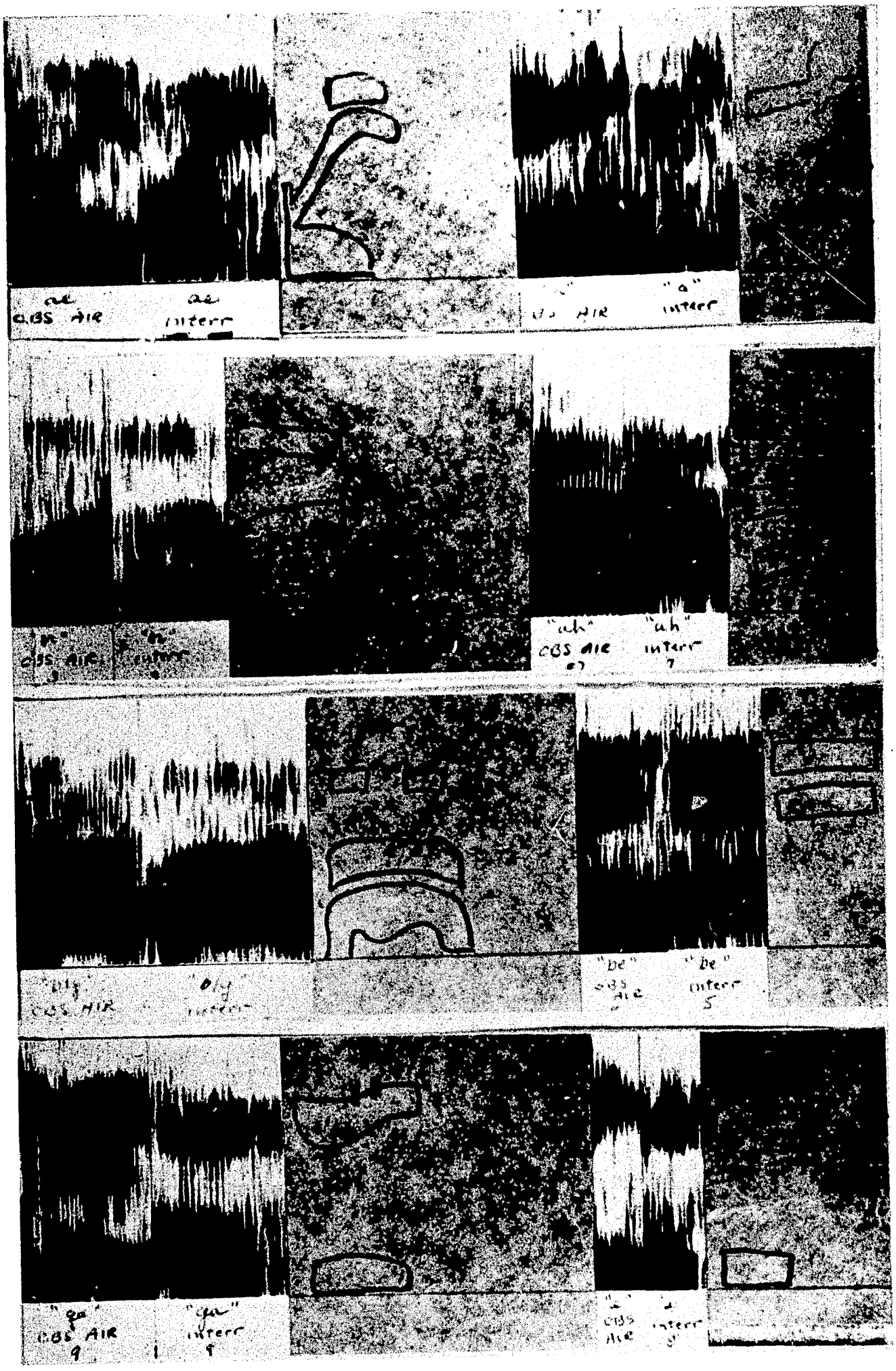
"ae"
CBS AIR
"ae"
interr
"a"
interr
"ah"
CBS AIR
"ah"
interr
"De"
CBS AIR
"De"
interr
"g"
CBS AIR
"g"
interr

Figure B-2