

On the comparison between identification and discrimination tests in speech perception

IRWIN POLLACK and DAVID PISONI
The University of Michigan, Ann Arbor, Mich. 48104

The conversion between identification and three forms of discrimination tests, based on the extreme assumption that discrimination performance in speech perception tests is determined solely by labeling in identification tests, is examined. Identical conversions are obtained for the two-interval same-different test, for the four-interval forced-choice test of pair similarity, and for the three-interval ABX test. Therefore, differences in discrimination performance among these tests are presumably due to their respective task requirements, not to differential prediction.

An extreme view of speech perception is that, under defined conditions, listeners can only "hear" speech-like materials in a linguistic mode (Liberman, 1970). For example, let us create a series of synthetic speech sounds that range perceptually from "bah" through "dah" to "gah." This extreme view holds that listeners can respond to these sounds only in terms of the encoded "bah, dah, gah" categorization. We label this view of speech perception as "extreme" not to discredit it, but rather to recognize that strong quantitative predictions can be made from this point of view without entertaining additional assumptions to achieve better prediction. The relationship of this strong prediction to the motor theory of speech perception has been enunciated in a series of papers from the Haskins Laboratories (Liberman, Harris, Hoffman, & Griffith, 1957; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; and especially Studdert-Kennedy, Liberman, Harris, & Cooper, 1970).

APPROACH

It is convenient to define two classes of tests: identification and discrimination. In identification tests, stimuli are presented one at a time. The task of the listener is to label the different sounds, usually into a set of categories defined by the E, e.g., into the sounds "bah, dah, and gah." In discrimination tests, several stimuli are presented and a variety of discrimination tasks may be defined. We consider three such tasks.

In the ABX discrimination test, a sequence of three sounds is presented: e.g., ABA or ABB, where A and B are stimuli employed in the identification tests. The first two stimuli, A and B, differ; the third is equally likely to be either A or B. The task of the listener is to indicate whether the third sound was more like the first or the second sound.

This particular test has been extensively employed in speech perception experiments at the Haskins

Laboratories. It possesses the excellent property that the E need not specify the dimension along which A and B differ, e.g., select the stimulus with the higher "pitch." The test is identical to the oddity problem employed in animal testing, except that the third item of the sequence cannot be the odd member. The test, however, has the disadvantage of placing a severe load on memory in that the third member of the sequence is to be compared both to the immediately preceding second member and to the remote first member. Thus, to solve

the discrimination problem, listeners may be forced to encode the stimuli in the manner predicted from the extreme view and consequently reveal the expected categorical discrimination functions. Conceivably, the listener could convert the ABX into a two-pair test by comparing the first and second sounds, then the second and third sounds. Os rarely report this strategy.

In the 4IAX test, a sequence of four sounds is presented: AA AB or AB AA. The task of the listener is to indicate whether the first pair was more alike than the second. This test has not been extensively employed by workers in speech perception. Indeed, its use by one of us (Pisoni, 1971) motivated the present search for conversion formulas, since speech discrimination tests showed consistent differences between the results of the ABX and 4IAX testing procedures. Both procedures have an advantage in that the basis for discrimination need not be specified by the E. The 4IAX procedure has additional merit in that the correct answer can be achieved by a pair-by-pair comparison, i.e., it is not necessary to compare the first and third stimuli, as is usually done in the ABX procedure.

Table 1
Derivation of Formulas

A. Notation						
		$b = P(b A)$	$b' = P(b B)$			
		$Q = bb + dd + gg$				
		$R = bb' + dd' + gg'$				
		$S = b'b' + d'd' + g'g'$				
		$1 - Q = bd + bg + db + dg + gb + gd$				
		$1 - R = bd' + bg' + db' + dg' + gb' + gd'$				
		$1 - S = b'd' + b'g' + d'b' + d'g' + g'b' + g'd'$				
B. Derivation						
Test	Type	Example	Correct		Guess	
			Probability	Example	Probability	Example
2IAX	AA	bb	Q			
	AB	bd'	(1 - R)			
	BA	d'b	(1 - R)			
	BB	b'b'	S			
4IAX	AA AB	bb bd'	Q(1 - R)	bb bb'	QR	(1 - Q)(1 - R)
	BB AB	b'b' bd'	S(1 - R)	b'b' dd'	SR	(1 - S)(1 - R)
ABX	ABA	bd'b		bb'b		
	ABB	bd'd'		bb'b'		
C. Formulas						
2IAX	$4P(C) = Q + S + 2(1 - R)$			} $2 + (i - i')^2$ Short Form		
4IAX	$4P(C) = 2Q(1 - R) + 2S(1 - R) + QR$ $+ (1 - Q)(1 - R) + SR + (1 - S)(1 - R)$ $= Q + S + 2(1 - R)$					
ABX	$4P(C) = Q + (1 - R) + S + (1 - R)$ $= Q + S + 2(1 - R)$					
3IFC	$6P(C) = 3(Q + S) + 2i[jk' + j'k'] - (ii' + i'i')$					
4IFC	$8P(C) = ii(3ij' + 3ik' + jj' + kk' + 2k'j + 2kj' + i)$ $+ i'i' (3i'j + 3i'k + j'j + k'k + 2j'k' + 2k'j + i')$					

In the 2IAX test, a sequence of two sounds is presented: AA or AB. The task of the listener is to indicate whether the two members sound the "same" or sound "different." The disadvantage of this test is that the criterion for "same" or "different" is under the control of the listener. As in the other tests, the task of the listener is to select the more likely event between the two available alternatives. The 2IAX test shares the advantage of the other procedures in that the basis for discrimination need not be specified by the E; it shares the advantage of the 4IAX procedure in that pairwise information is sufficient to define a correct response.

Appropriate counterbalancing of the order of presentation is required for all three tests; e.g., in the ABX tests, we should employ an equal number of ABA, ABB, BAA, and BAB tests.

DERIVATIONS

We assume that in a discrimination test two sounds will be nondiscriminable whenever they are assigned the same label in an identification test and will be discriminated only if assigned different labels.

Consider first the 2IAX test. If each of the two sounds of a pair of sounds is assigned the same label in identification tests, the assumption is that the listener will respond "same" in the discrimination tests; if each pair of sounds is assigned a different label in identification tests, we will assume the listener will respond "different" in discrimination tests. The single assumption does not admit to modification of the criterion for responding "same" or "different" and, therefore, is not powerful enough to predict the full complexity of the same-different test. Nevertheless, it may be satisfactory for describing overall error proportions, assuming a symmetrical payoff matrix and nonextreme "same-different" proportions.

Table 1 illustrates the derivation. The proportion of b, d, or g responses in the identification test to the A stimulus is represented as $b|A$, $d|A$, or $g|A$, respectively. The proportion of b, d, or g responses in the identification test to the B stimulus is represented as $b|B$, $d|B$, and $g|B$, respectively. The next set of terms under "Notation" provides a pairwise classification of terms which will be useful in the derivation.

There are four types of trials in the 2IAX tests, as shown, each with probability 0.25. Examples of AA trials are: bb, dd, and gg. The

contribution to a correct "same" response, given by an AA trial, is $(bb + dd + gg)$ or Q. The contribution to a correct "different" response, given by an AB trial, is $(bd' + bg' + db' + dg' + gb' + gd')$ or $(1 - R)$. The contribution to a correct "different" response, given by a BA trial, is, assuming no time order effects, $(1 - R)$. And, the contribution to a correct "same" response, given by a BB trial, is $(b'b' + d'd' + g'g')$ or S. The final proportion of correct trials, $P(C)$, is then:

$$4P(C) = Q + S + 2(1 - R)$$

The same basic procedure is employed for evaluating the 4IAX and ABX procedures. Upon AA-AB trials, for example, a correct discrimination response in the 4IAX test is registered if the label assigned to each A stimulus of the first pair is in agreement and if the labels assigned to each stimulus member of the second pair are in disagreement. [The following notation is employed in the examples: if a specific mismatch is demanded, the typical example employs a "g"; if any mismatch can be used, the typical example employs a "d"; if any response can be used, the typical example employs a "b".] When possible quadruplets are enumerated, the contribution to correct responses is $[bb(1 - R) + dd(1 - R) + gg(1 - R)]$ or $Q(1 - R)$. Similarly, upon BB-AB trials, a correct selection is made if the label assigned to each B stimulus of the first pair is in agreement and the labels assigned to the stimulus members of the second pair are in disagreement. There are two ways to arrive at a correct response by guessing: when the labels agree both within the first pair and within the second pair, and when the labels disagree both within the first pair and within the second pair. If guessing is random, the first way contributes $(QR)/2$ or $(SR)/2$, whereas the second way contributes $(1 - Q)(1 - R)/2$ or $(1 - S)(1 - R)/2$. Again, the same formula is obtained, as in the 2IAX tests, for $P(C)$.

The ABX procedure has been examined previously in detail (Liberman et al, 1957). A correct response in an ABA trial is obtained when the labels assigned to the first and third A stimuli are in agreement and, in turn, differ from that assigned to the B stimulus. Now there are three ways to arrive at a correct response through guessing: when all three stimuli give rise to the same label; when the first two stimuli yield the same label but differ from the third; and when the labels differ to each of the three stimuli. The collation of

trials into the present notation follows: the weighting to correct responses of the correct examples is twice that of the guess examples. We, therefore, duplicate the correct examples: one set combines with the first guess example to yield $(bd'b + bg'b + bb'b)$ or $bb(b' + d' + g')$ or bb , which leads to the Q term for the ABA trials and to the S term for the ABB trials; one set combines with the second and third guess examples to yield $b(d'b + g'b + b'd + b'g + d'g + g'd)$ or $b(1 - R)$, which leads to the $(1 - R)$ term for ABA tests, and $b'(1 - R)$, which leads to the $(1 - R)$ term for ABB tests.

The three tests predict the same conversion formula. An alternative short form is also presented in Table 1, paralleling that of Liberman et al (1957), where i represents each of the three alternatives in turn, j represents either of the remaining alternatives, and k represents the remaining alternative. Thus $(i - i')^2 + (b - b')^2 + (d - d')^2 + (g - g')^2$: We have also explored the conversions for the 3IFC and 4IFC oddity tests. Since their chance response levels will differ from that of the two-alternative tests, their conversions, also shown in Table 1, cannot be the same. (Across the two major terms of 4IFC, i and i' can be assigned independently.) To date, we have been unable to simplify the conversions for the oddity tests in terms of R, S, and Q alone.

Finally, we note that differences in obtained discrimination scores with the three initial procedures do not arise from differential prediction by the labeling hypothesis. Rather, differences in the obtained discrimination scores across the procedures are presumably to be sought in the different task requirements embodied within the individual tests.

REFERENCES

- LIBERMAN, A. M. Some characteristics of perception in the speech mode. In D. A. Hamburg (Ed.), *Perception and its disorders: Proceedings of A.R.N.M.D.* Baltimore: Williams & Wilkins, 1970. Pp. 238-254.
- LIBERMAN, A. M., COOPER, F. S., SHANKWEILER, D. P., & STUDDERT-KENNEDY, M. Perception of the speech code. *Psychological Review*, 1967, 74, 431-461.
- LIBERMAN, A. M., HARRIS, K. S., HOFFMAN, H. S., & GRIFFITH, B. C. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 1957, 54, 358-368.
- PISONI, D. B. On the nature of categorical perception. Unpublished doctoral dissertation, Department of Psychology, University of Michigan, September 1971.
- STUDDERT-KENNEDY, M., LIBERMAN, A. M., HARRIS, K. S., & COOPER, F. S. Motor theory of speech perception: A reply to Lane's critical review. *Psychological Review*, 1970, 77, 234-249.