Discrimination of Three-Formant Nasal-Vowel Syllables. (III)

Erica Garcia†

Haskins Laboratories, New York City

In the Status Report on Speech Research for Jan-March 1967 (SR-9) we discussed the identification of synthetic nasal-vowel syllables based on three-formant vowel stimuli.[1] It was found in that experiment that labelling of synthetic nasals -- m, n, and ŋ, each presented with the vowel -- was equally sharp in CV and in VC position, and that "good" labellers showed as abrupt a cross-over from one nasal category to the other as had earlier been observed in the case of voiced stops.[2]

The similarity of labelling functions for stops and nasals might lead one to expect that the latter too would be perceived "categorically," which is in fact what one would anticipate on the basis of the "motor" theory of speech perception, since

---

† Also, Columbia University

1. That study was a continuation of an earlier experiment with two-formant vowel stimuli, included in SR-7/8 under the title: "The identification and discrimination of synthetic nasals."

2. These data are presented in A.M. Liberman, K.S. Harris, H.S. Hoffman, and B.C. Griffith, "The discrimination of speech sounds within and across phoneme boundaries." J. Exptl. Psychol. 54, 358-368 (1957); their implications for the "motor theory" of speech perception are discussed, most recently, in A.M. Liberman, F.S. Cooper, D.P. Shankweiler, and M. Studdert-Kennedy, "Perception of the Speech Code." Psychol. Review 74, 431-461 (1967) and also in SR-9 (1967).

nasals are very like voiced stops from an articulatory point of view, even though they share some acoustic properties with vowels, e.g. formant structure. In particular, the difference between m, n and ŋ is exactly the same as that between b, d and g.

We shall now report on the result of the discrimination tests that followed the labelling runs. At least 28 judgments were made on each paired comparison. Each pair was presented in the form of ABX triads of one, two and three step distances along the acoustic continuum being investigated (change of second and third formant transitions.) The data are presented in the usual form, i.e., the graphs contrast the observed or actual discrimination function with the one predicted on the basis of the subjects' labelling performance. In Figs. 1A and 1B we give the curves for two of the three tests: III, involving the continuum from m to n, and IV, for ŋ and n, in both CV and VC positions. The following representative subjects have been selected:[3]

VH: the best labeller of the group.

PN: the worst labeller of the group.

EC: a non-labeller, whose "predicted" curve runs close to 50% for all tests, and who could be regarded as a "non speech" control on the discrimination tests.

Overall Good: the overall results for all "good" labellers, as defined in our SR-9 report.

Overall: the overall results for the entire group, i.e. all those capable of labelling. Here only EC was excluded.

---

3. In the SR-9 report we presented the labelling curves for VH, PN and Overall.

The following observations are suggested by the data underlying Fig. 1A and Fig. 1B:

1. The fit between actual and predicted curves, as well as the general level of discrimination, is considerably better than that obtained with two-formant syllables. (See our report in SR-7/8.) This strongly suggests that the "categoricalness" of perception may be influenced, at least in part, by the naturalness of the stimuli presented.

2. As in our earlier experiment, initial vs. final position appears to play no role in the discriminability of nasals: the fit between the actual and the predicted functions, and the level of both, are approximately the same for both CV and VC tests. In this respect, nasals appear to differ strikingly from stops. Though no study has yet been made of the identification and discrimination of stops in final position, preliminary work indicates clearly that for stops synthesized by means of transitions alone, identification is much more difficult in final than in initial position.[4]

3. There is an interesting difference between tests III and IV. In III, the actual discrimination peaks occur precisely where they are predicted on the basis of the labelling performance of the subjects. In test IV, on the other hand, particularly for the three-step pairs, we observe a displacement of the observed discrimination curve with

---

4. Cf. A.M. Liberman, P.C. Delattre, F.S. Cooper, and L.J. Gerstman, "The role of consonant-vowel transitions in the perception of the stop and nasal consonants," Psychol. Monographs: Gen. & Applied 68, No. 8 (1954) esp. p. 12.

DISCRIMINATION OF m to n CONTINUM

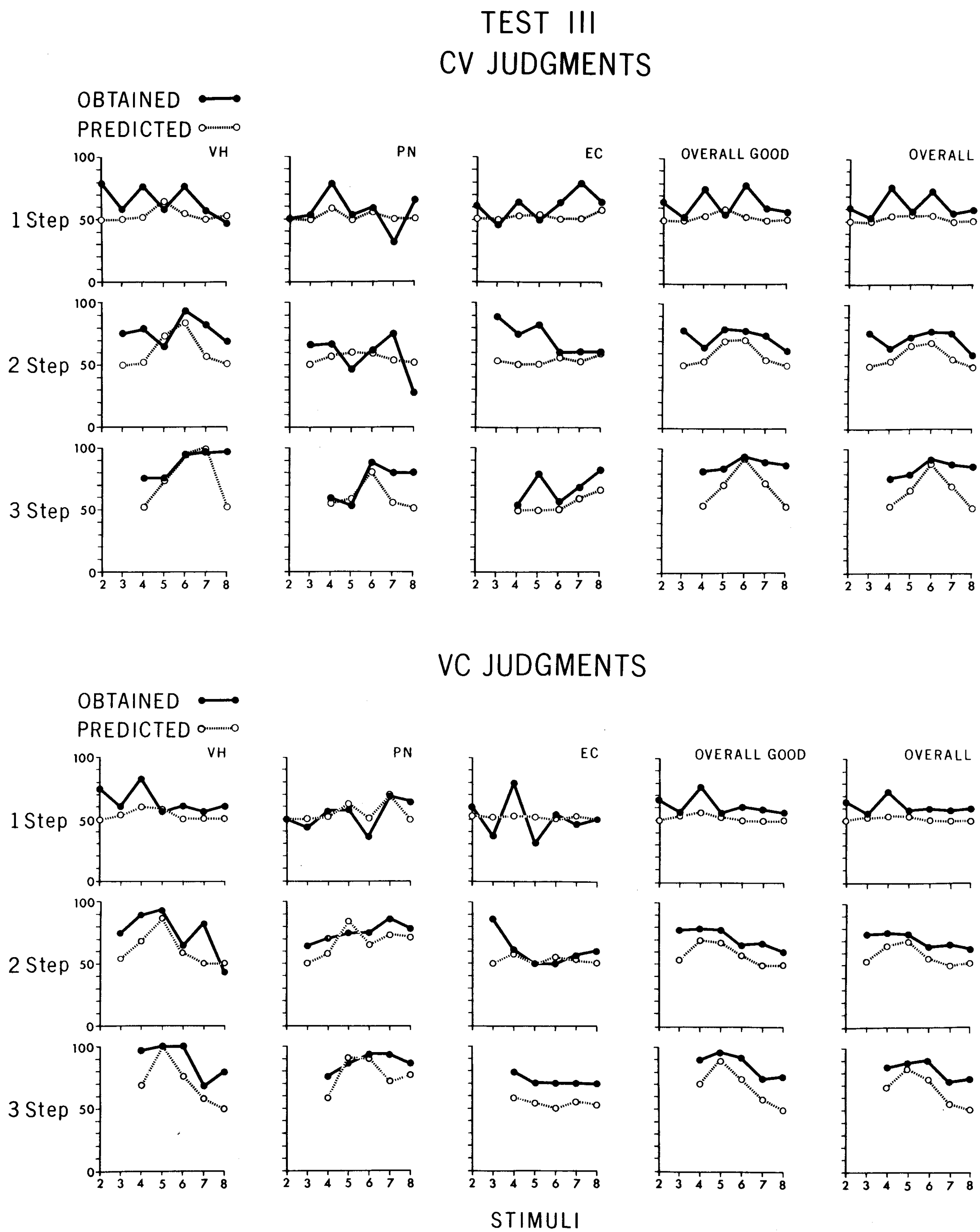
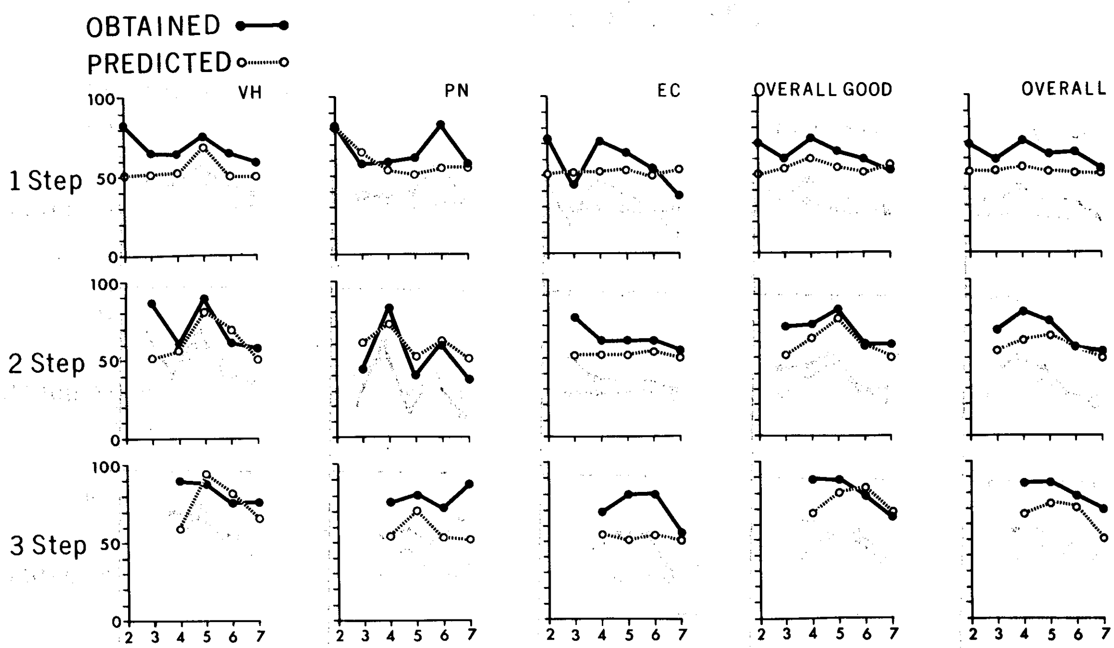


FIG. 1A

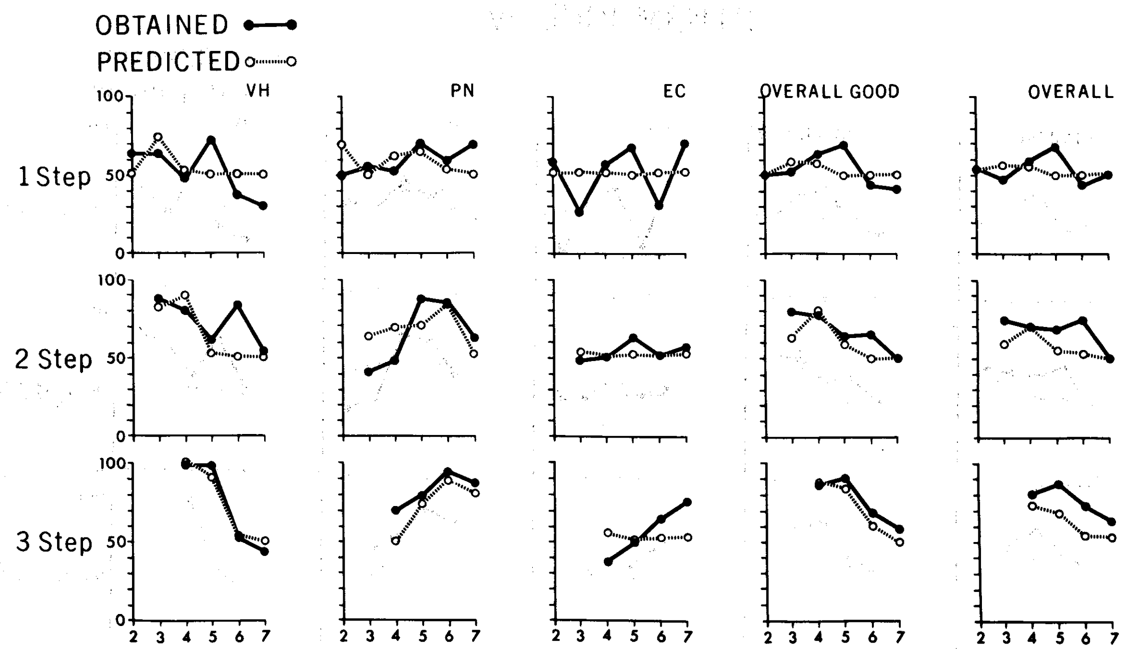
DISCRIMINATION OF ŋ to n CONTINUM

TEST IV

CV JUDGMENTS



VC JUDGMENTS



STIMULI

FIG. 1B

respect to the "predicted" one: to the left, in CV stimuli, but to the right, for the VC tests. The reason for this shift, and for the discrepancy between CV and VC tests is unknown.

4. The actual discrimination functions of the non-labeller EC and of the poor labeller PN follow, in general, the pattern established by good labellers. Moreover, in the case of PN in test IV, the two-step and three-step curves are in remarkably close agreement with the predicted functions On the other hand, in thos cases where the poor and non-labellers do not coincide with the overall functions of the good labellers, they generally also fail to agree among themselves. This is particularly clear for test III.

The disagreement among the poor labellers is significant for the following reason: these subjects constitute our only hope of obtaining a non-speech control for the discrimination of nasals. It is well-nigh impossible to use the variables of this experiment to construct a non-speech set of stimuli that might serve as a control on the inherent difficulty of the different paired comparisons along the continuum. We can consequently only hope that the poor labellers, unable to hear the synthetic stimuli as speech, will have to depend exclusively on their physical properties, rather than on their linguistic associations, in order to identify X with either A or B. They can thus be expected to reveal whether any pair of stimuli on the continuum is inherently more discriminable than another. We find, however, that the poor labellers were unable to discriminate among the stimuli in these tests in any consistent way, save for the cases when they agreed with their own predicted functions or with the behavior of the Overall Good group. Our "non-speech" control thus suggests that there is no one point where the acoustic continuum "breaks" for non-linguistic reasons, and at which there might occur a "peaking" of the discrimination functions unrelated to the subjects' ability to label the stimuli.

5. The functions of the Overall Good and Overall groups are remarkably similar. The shapes of both predicted and actual discrimination curves are the same for both groups; as is natural, the predicted curves rise higher for the Overall Good. This only confirms our conclusions in the preceding paragraph: the poor labellers do not discriminate in a consistently different way from the good ones, but depart randomly from the "good" pattern. They can thus lower the general level of the observed discrimination function, but not alter its shape. What is particularly worthy of notice is that the <u>fit</u> between predicted and observed discrimination does not improve for the good labellers alone as opposed to the entire group. The observed function follows the predicted, but at a distance; in other words, actual discrimination throughout the continuum is consistently better than would be anticipated on the assumption that the subjects discriminate <u>only</u> as well as they label. This is particularly clear in test III. In other words, nasals do not appear to be categorically perceived, in the strict sense of "categorically."

The lack of close fit between the two discrimination functions sets nasals apart from stops, without however grouping them with vowels, where the lack of fit between the curves is much greater. For purposes of comparison, we reproduce the

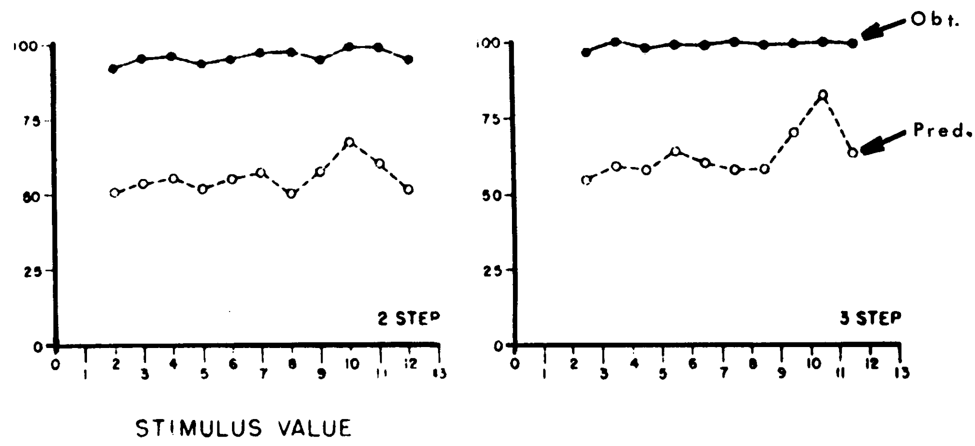


FIG. 2A

Group obtained and predicted discrimination for 2- and 3-step differences among the synthetic vowel stimuli.

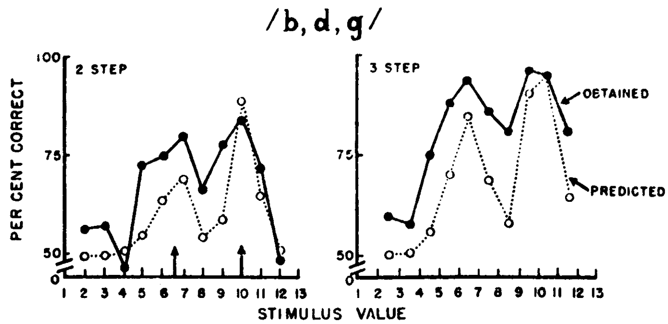


FIG. 2B

Group obtained and predicted discrimination functions for the /b,d,g/ stimuli at 2 and 3 steps. [Reproduced from Eimas (1962), Fig. 4.] References may be found in SR 7/8.

the relevant graphs for synthetic vowel and stop discrimination in Fig. 2A-2B.

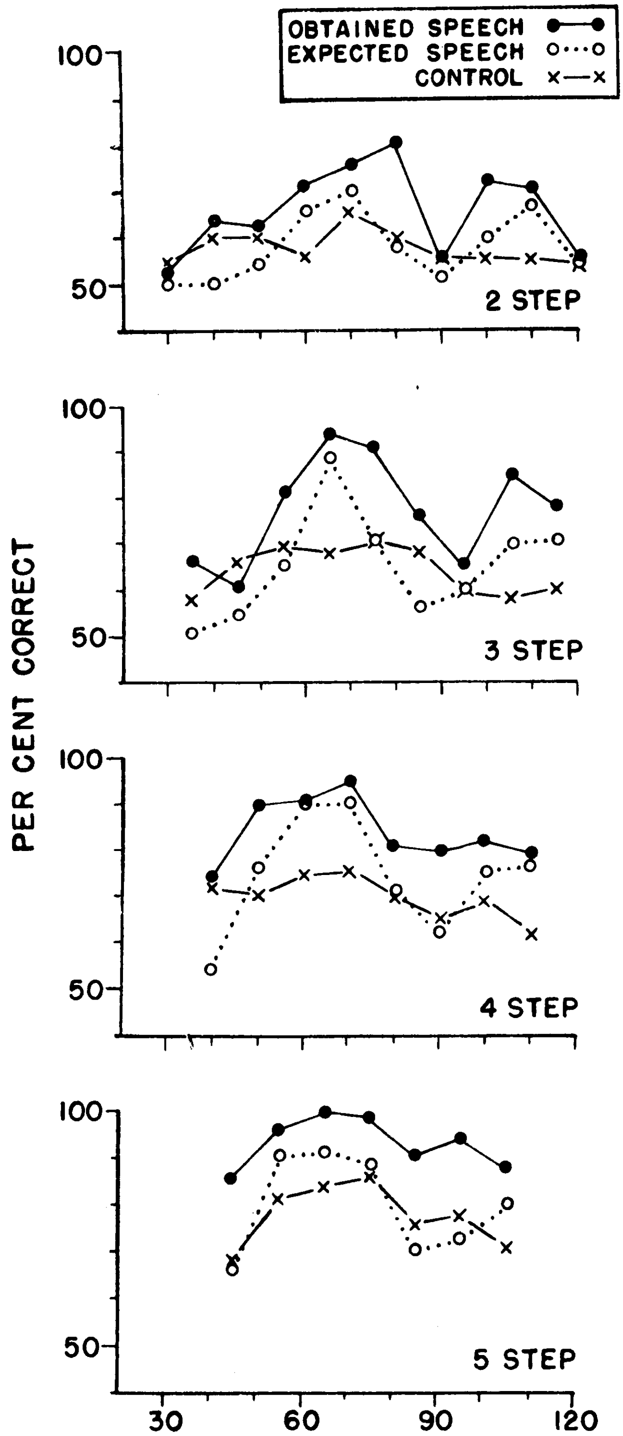
Quite clearly, the curves for nasals fall between the extremes of vowel discrimination and /b,d,g/ discrimination as one would have anticipated on the basis of their dual resonant-"stop" characteristics. It does however raise the question as to whether a literal interpretation of the motor theory, which must group nasals with stops, is compatible with out data.

It should be pointed out that a degree of fit between predicted and observed discrimination function that resembles very closely that of our curves was obtained by Liberman et al in a study of the perception of durations of silence.[5] Here too, the actual discrimination follows the general pattern of the predicted curve, but at a considerably higher level. In Fig. 3 we reproduce the relevant graphs.

These cases of intermediate fit raise the question as to how categorically categorical perception can be defined. It is not clear why the voiced-voiceless distinction (synthesized by means of different durations of silence in the study by Liberman et al mentioned in footnote 5) should be less categorically perceived than that among b, d, and g, since in both cases there is an obvious articulatory discontinuity. Moreover, the articulatory discontinuity among the nasals is exactly the same as that among the stops, yet the fit between discrimination functions is the same as that for the voiced-voiceless opposition. There

5. A.M. Liberman, K.S. Harris, P.D. Eimas, L. Lisker and J. Bastian, "An effect of learning on speech perception: the discrimination of durations of silence with and without phonemic significance," Language and Speech 4, 175-195 (1961).

FIG. 3



Obtained and expected speech discrimination functions for the non-speech control stimuli. (See footnote 5.)

is not a clear explanation of these facts on the basis of how the sounds are produced by human speakers. However, since the naturalness of the stimuli appears to play a considerable role in promoting "categorical" perception, there is the unwelcome possibility that all cases of lack of fit between observed and predicted functions, including perhaps even that of vowels, may simply be due to the inadequacy of the acoustic stimuli rather than to articulatory properties of the speech sounds synthesized.