74

PAPERS FROM THE EIGHTEENTH REGIONAL MEETING CHICAGO LINGUISTIC SOCIETY

April 15–16, 1982

Edited by

Kevin Tuite
Robinson Schneider
Robert Chametzky

CHICAGO LINGUISTICS SOCIETY
THE UNIVERSITY OF CHICAGO
CLASSICS 314A
1050 E. 59th STREET
CHICAGO, ILLINOIS 60637

TABLE OF CONTENTS

Anthony Aristar and Helen Dry: The origin of backgrounding tenses in English 1
Peter C. Bjarkman: Process versus feature analysis and a notion of linguistically "closest" sounds 14
Frank Roberts Brandon: Q-Float and Conjunction Reduction as evidence for NEG-Placement 29
Allen C. Browne: Inductive inference and pragmatic explanation 40
Greg Carlson and Michael Tanenhaus: Some preliminaries to psycholinguistics 48
Robert Channon: On the English pronoun Ø 61
Amy Chukerman (see Gross and Chukerman)
Amy Dahlstrom: A functional analysis of switch-reference in Lakhota discourse 72
Bill J. Darden: Dissimilative glide insertion in West Greenlandic 82
Michael Dobrovolsky: Some Turkic dentals 87
Helen Dry (see Aristar and Dry)
W. Neil Elliott: A case of learning Case 95
James Emil Flege: English speakers learn to suppress stop devoicing 111
Robert Allen Fox: Adaptation level and speaker quality in vowel identification 123
Samuel Ethan Fox: Autosegmental phonology and Semitic 131
Steven Franks: Is there a Pro-drop parameter for Slavic? 140
Steven Franks and Richard House: *Genitive themes in Russian* 156
Mary Fritz (see McClure, Saville-Troike and Fritz)
Kurt Godden: Causing ambiguity 169
Haia Gross and Amy Chukerman: Functional morphophonemics: West Greenlandic Eskimo 176
Luise H. Hathaway: Style shifting as metaphorical change in point of view 185
Randall Hendrick: Eskimo feet 191
Richard House (see Franks and House)
Grover Hudson: Reply to Darden 1979 202
Larry G. Hutchinson: Sentential and subsentential coordination in linguistics and logic 209
Pauline Jacobson: Visser revisited 218
Richard Douglas Jehn: On the notion 'inherent segment duration' in phonetic timing research 244

Adaptation level and speaker quality in vowel identification*

Robert Allen Fox
The Ohio State University

INTRODUCTION

As has often been noted in the literature, the identification of a particular vowel sound can be significantly affected by the surrounding phonetic context. For example, Ladefoged & Broadbent (1957) demonstrated that the formant pattern of a precursor phrase could affect the category into which listeners placed a subsequent ambiguous speech token. Similar results have been obtained by Eimas (1963), Fry et al. (1962), and Thompson & Hollien (1970). In general this contextual influence is one of contrast rather than assimilation.

In the last decade, most of the information concerning this contextual influence on phonetic decisions has been obtained using either the selective adaptation paradigm or, more recently, anchoring procedures. Selective adaptation involves presenting a subject with repeated occurrences of one speech sound and then testing his identification of a set of stimuli (usually a speech continuum). These adapted identifications are then compared with the subject's unadapted, baseline identification of the same stimuli. The most common result of such experiments is to show that the adaptation condition shifts the category boundary of the test stimuli toward that of the adaptor. Anchoring procedures are similar to selective adaptation in that they involve presenting many occurrences of one stimulus to subjects and testing how their identifications on other test items are affected, but differ in that these repeated stimuli are randomly intermixed among the test items rather than presented in a concentrated, repetitive manner before the identification test proper. Results from the anchoring condition are compared to an equiprobable control condition wherein subjects hear an equal number of all test items. In both of these procedures, contrast effects are found such that more test items are identified as being from a different phonetic category than that of the more frequently occurring category.

There are several competing theories explaining the results of such experiments including reference to response bias (Parducci, 1974), feature detector fatigue (see Ades, 1976), and adaptation level (Sawusch & Nusbaum, 1979). However, recent work done using anchoring procedures and critical re-evaluations of the selective adaption literature have argued strongly that the effects of context upon vowel identification can best be explained in terms of a version of Helson's (1964) adaptation level theory which claims that vowel quality is determined on the basis of three factors: (1) the acoustic nature of the vowel itself, (2) the immediate phonetic context, and (3) the auditory background or adaptation level (Restle, 1978). The auditory background is an appropriately weighted measure of the acoustic characteristics of the stimuli previously presented. If there are repeated occurrences of a particular stimulus (i.e., an anchor or adaptor), then an auditory ground is established to which the stimulus has contributed a disproportionate amount of auditory information. Ambiguous boundary stimuli later presented for identification will contrast with this ground and will be placed in a different category from the more frequently occurring stimulus item (Simon & Studdert-Kennedy, 1978). However, what is the precise nature of the auditory information being held in either immediate memory or the auditory background? Is this auditory information retained in something like Crowder's (1973) PAS (precategorial acoustic store) and thus stored in a veridical fashion (as in a taped recording) or are only certain auditory features (e.g., normalized formant-patterns) maintained in a form similar to that found in Massaro's (1975) synthesized auditory memory?

To address this issue, an anchoring experiment was designed to discover the extent to which listener's identifications of a set of vowel stimuli would be differentially shifted by anchors which, although of the same phonetic value, would differ in terms of F0 or absolute formant frequencies, paralleling the speaker qualities variations one finds in vowels of the same phonetic quality.

EXPERIMENTAL DESIGN

The experiment to be described investigated the effect which three different versions of an [hɪd] anchor had upon the identification of a 7-step [hɪd]--[hɛd] continuum. The anchors included: (1) the [hɪd] endpoint of the test continuum (Anchor 1); (2) a stimulus token identical to Anchor 1 except in terms of F0 contour (Anchor 2); and (3) a stimulus token having different absolute formant frequencies to Anchor 1 but similar in terms of relative formant pattern F0 (Anchor 3).

Stimuli: Seven steady-state vowels were synthesized in an [h_d] context. The second and third formants varied in approximately equal steps while the first formant varied in logarithmic steps. The stimuli were synthesized using the Klatt synthesis program[1] and the relevant acoustic parameters are shown in Table 1. Each stimulus token was 420 msec in duration.

TABLE 1. Formant frequencies for the /h vowel d/ stimuli.

Stimulus	F1	F2	F3
1	400	1800	2570
2	422	1780	2558
3	443	1760	2547
4	465	1740	2535
5	487	1720	2523
6	508	1700	2511
7	530	1680	2500

The fundamental frequency (F0) for each test vowel began at 220 Hz and fell linearly to 190 Hz after 360 msec at which level it remained until the end of the token. The formant transitions to the final [d] were 30 msec in duration and went from the formant values specified in Table 1 to 200, 1600, and 2600 Hz for F1, F2, F3, respectively. The frequencies of F4, F5 and F6 were set at 3300, 3850, and 4000 Hz, respectively. Shown in Table 2 are the relevant acoustic parameters of the three anchor stimuli.

TABLE 2. Acoustic parameters for the three anchor stimuli.

Acoustic Parameter	Anchor 1	Anchor 2	Anchor 3	
F1	400	400	400	Hz
F2	1800	1800	1800	Hz
F3	2570	2570	3000	Hz
F0	220-190	130-110	220-190	Hz
Duration	420	420	420	Msec

Note that Anchor 1 represents the [hɩd] endpoint of the 7-step test continuum (Stimulus 1). Anchor 2 differs from Anchor 1 only in terms of F0, that is, Anchor 2 begins at 130 Hz and falls linearly to 100 Hz after 360 msec where it remains steady until the end of the token. Both have the same 30 msec formant transitions into [d]. Anchor 3 has an F0 contour identical to Anchor 1 but the formant frequencies are much higher in terms of absolute values (though the relative formant structure is similar). The formant transitions into [d] for Anchor 3 were 400, 2000 and 2900 Hz, for F1, F2 and F3, respectively. In terms of reflecting normal human voices, Anchor 1 and the test continuum represent acoustic values characteristic of a male speaker, but with a high F0. Anchor 2 also represents a male vocal tract, but with a much lower F0. Anchor 3 reflects acoustic parameter values more commonly associated with a female vocal tract.

The stimuli were converted into analog form using a PDP 11/23 computer in the Linguistic Laboratory at The Ohio State University and recorded on audio tape using a BIC T-2M stereo cassette tape deck. Four different stimulus tapes were made: one control tape and three anchor tapes. The control tape contained 10 different examples of each of the 7 different test tokens in 2 different random orders, yielding a total of 140 stimuli plus a set of practice items. One anchor tape contained 10 different examples of the 7 different test tokens plus 30 additional examples of the [hɩd] endpoint serving as Anchor 1 in two different random orders, yielding 200 different stimuli. The other two anchor tapes contained 10 examples of the 7 different test tokens plus 30 examples of either Anchor 2 or Anchor 3 in two different random orders. The anchor tapes thus contain 20 examples of each of the test tokens plus at least 3 times as many of the anchor stimuli as any other single token. All stimuli were recorded with a 4 sec inter-stimulus interval.

Procedure. Baseline identifications using the equiprobable control tape were obtained from each subject tested individually before the anchoring condition. Subjects were required to identify each stimulus either as *hid* or *head* by circling the appropriate word on the answer sheet. In addition, subjects were required to rate their identification response indicating how sure they were that they had identified the vowel correctly. A 4-point rating scale was used with a "1" indicating that the subject was positive the identification was correct and a "4" indicating that the subject was guessing. These identification ratings were then converted into an 8-point scale with a 1 indicating that the subject was positive the token was *hid* and an 8 indicating the subject was positive the token was *head*. Following presentation of the baseline tape, each subject heard one of the three anchor tapes. As before subjects were required to both identify the stimulus and rate their responses. Subjects were not given any information concerning possible differences between the control tape and the anchor tapes.

Subjects. The subjects were divided up into three groups of eight subjects each. They were run individually in an anechoic chamber, and the entire test took approximately 35 minutes to complete. The stimulus tapes were reproduced on a BIC T-2M tape deck and presented binaurally to subjects at a comfortable intensity level through Realistic Nova-PRO earphones. The subjects were informed only that they would hear a series of synthetic speech tokens that would sound like *hid* or *head* and were to respond in their test booklets as described above.

RESULTS AND DISCUSSION

The results[2] from the Anchor 1 group are shown in Fig. 1. Note that there is a significant [ɩ]/[ɛ] category boundary shift toward [ɩ] ($t(7)=8.24$, $p < .001$, tailed). The mean baseline boundary was 3.86 which shifted to 2.87 in the anchoring condition. This result is consistent with Sawusch & Nusbaum (1979) who demonstrated the effectiveness of the vowel [ɩ] in shifting the [ɩ]/[ɛ] boundary in an anchoring procedure and the study by Morse et al. (1976) which found a marginally significant shrinkage of the [ɩ] responses following adaptation with an [ɩ] adaptor. It should be noted that finding equivalent results with both anchoring and adaptation procedures do not support the feature detector explanation of such effects since the occurrences of the anchor vowel appear at widely different intervals and make fatigue of putative detectors quite unlikely[3].

The results for the Anchor 2 group are shown in Fig. 2. Here again there is a significant shift of the category boundary toward [ɩ] ($t(7)=3.15$, $p < .01$, 1-tailed). The mean baseline boundary was 3.80 which shifted to 3.35 in the anchor condition. Comparing Figs. 1 & 2, we find that the shift using Anchor 2 is

Fig. 1. Anchor 1 group.

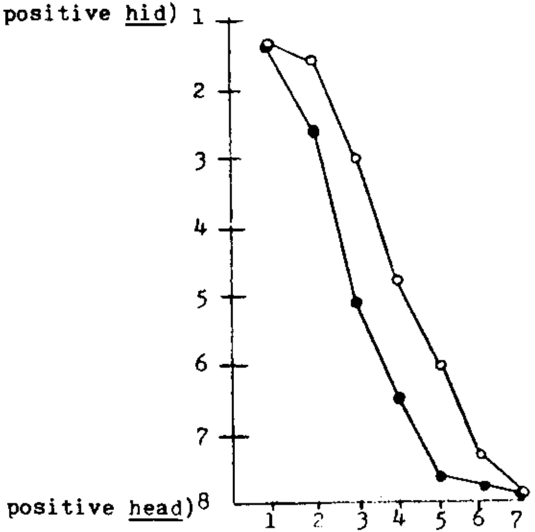


Fig. 2. Anchor 2 group.

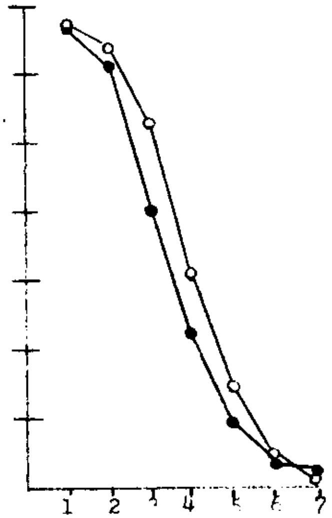
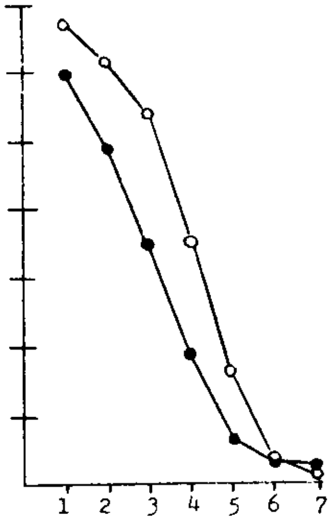


Fig. 3. Anchor 3 group.



Note: Baseline (open circles) and anchored (closed circles) rating functions for the three different experimental groups.

only half that found using Anchor 1. These data suggest that we must take into account more information than the absolute formant frequency values in accounting for the contrast effect. In particular, one cannot explain category boundary shifts in terms of a simple comparison between the F2-F1 ratios among the anchors and the test stimuli as did Thompson & Hollien (1970) for immediate vowel context effects since there is no formant frequency difference between Anchors 1 and 2.

As is well known, a particular vowel with a given formant pattern may be perceived as two different phonetic qualities depending on the F0 (Miller, 1953; Slawson, 1968; Fujisaki & Kawashima, 1968). This difference is explicable if we assume that listeners normally rescale the formant structure of a vowel to eliminate speaker quality variations due to vocal tract length differences and then identify a vowel on the basis of the rescaled formant structure. Note that the first formant of [ι] for a female speaker may be on the order of 10% higher than that of a male speaker (Fant, 1973). Since F0 is negatively correlated with the length of a speaker's vocal tract, if a listener utilizes F0 to estimate vocal tract length then, ceteris paribus, the perceived phonetic quality of given formant patterns will differ as a function of fundamental frequency. Turning back to the Anchor 2 results, if we assume that listeners rescale F1 of the anchor on the basis of F0, then lowering the fundamental will produce a relatively higher rescaled F1 relative to Anchor 1 and the test series. Since F1 is inversely proportional to vowel height, we find that a normalized Anchor 2 would be closer to test stimulus 2 or 3 than test stimulus 1 (which is Anchor 1). We now can explain the differential effects of Anchor 1 vs. Anchor 2. As Sawusch & Pisoni (1976) demonstrated in a selective adaptation experiment, the perceived acoustic attributes of an adaptor stimulus determines the direction and magnitude of the adaptation effects rather than the label of that stimulus. Therefore, despite the fact that Anchor 2 was labeled [hιd] over 95% of the time, its normalized formant pattern produced a lower vowel than presented in the Anchor 1 condition and therefore resulted in a smaller boundary shift.

The Anchor 3 results are presented in Fig. 3 and again show a significant boundary shift in the anchoring condition ($t(7)=7.83$, $p < .001$, 1-tailed). The shift in the mean category boundary from 3.99 in the baseline condition to 2.81 in the anchor condition was slightly greater than that found using Anchor 1, but this difference was not significant ($t(14)=.983$, n.s.). However, comparing the results using Anchor 3 with those using Anchor 2 we again find that the change in phoneme boundary is significantly different. Utilizing the formant normalization explanation suggested for Anchor 2 results, we can also explain the Anchor 3 results--in particular why formant values so different from Anchor 1 will produce equivalent category boundary shifts. As noted in Fujisaki & Kawashima (1968) (and discussed in Massaro, 1975) both F0 and F3 can function as reference levels in the rescaling of F1 and F2 to eliminate vocal tract length differences. Although Anchor 3 does not differ from Anchor 1 in terms of F0, it does differ in terms of having a much higher F3. If listeners rescale the first two formants by assuming it has been produced by a shorter vocal tract then the formants may be rescaled according to the scaling factors suggested in Fant (1973). Fant suggests that the scaling factor difference (kn) for [ɪ] is 10 for F1 and 24.5 for F2. Calculating the scaling factor differences on the basis of Anchors 1 and 3 we find factors of 12.5 for F1 and 27.8 for F2. If we actually rescale Anchor 3 relative to the test series we find that the normalized F1 and F2 values are very similar to Anchor 1. Thus, if the contrast effect is due to auditory contrast, it would again argue that this auditory information has undergone normalization.

It is clear that more work is needed in this area and the author is currently engaged in a large project utilizing test stimulus series based on both Anchors 2 and 3 using both [hɪd] and [hɛd] anchors. One particular concern will be in determining just how many levels of contrast are operative. For example, in the Anchor 3 results one finds that in addition to the boundary shift, the identification of the [hɪd] endpoint itself was affected, a significant 'within-category' shift. This particular effect could possibly be attributed to pure auditory contrast. Finally, these results provide another bit of evidence suggesting that some process like formant normalization is psychologically real.

Footnotes

*I would like to thank Jean Godby for reading this paper for me at the conference. This research was supported by a Small University Research Grant through the College of Humanities.

[1] The original Klatt programs (including the executive program) were kindly provided by Diane Kewley-Port and the Psychology Laboratory at Indiana University. These programs were later greatly modified by Takao Gunji to run on our PDP 11/23 system.

[2] All results are presented in terms of the converted 8-point scale.

[3] Another strong argument against the response bias explanation is that a different group of subjects were run on the same tapes (control tape and one of the anchor tapes) but were explicitly warned that for Anchor 1 many more examples of hid would appear on the tape and for Anchors 2 & 3 that completely different tokens (which were perceptibly different) from the test items would appear. Although they were required to identify tense anchors, in each case they were told to not let the additional items affect their identification of the test stimuli. These results (to be presented elsewhere) were not significantly different from those presented here, even with Anchors 2 & 3. These results are consistent with Sawusch & Nusbaum (1979).

References

Ades, A. E. (1976) "Adapting the property detectors for speech perception," in R. J. Wales & E. Walker (eds.) NEW APPROACHES TO LANGUAGE MECHANISMS. Amsterdam: North-Holland.

Crowder, R. G. (1973) "Representation of speech sounds in precategorial acoustic storage," JOURNAL OF VERBAL LEARNING & VERBAL BEHAVIOR, 10:587-590.

Eimas, P. D. (1963) "The relation between identification and discrimination along speech and non-speech continua. LANGUAGE & SPEECH 6:206-217.

Fant, G. (1973) SPEECH SOUNDS AND FEATURES, Chapt. 4. MIT Press.

Fry, D., Abramson, A., Eimas, P. D., & Liberman, A. M. (1962) "The identification and discrimination of synthetic vowels," LANGUAGE & SPEECH 5:171-189.

Fujisaki, H. & Kawashima, T. (1968) "The roles of pitch and higher formants in the perception of vowels," IEEE TRANSACTIONS ON AUDIO AND ELECTROACOUSTICS, AU-16:73-77.

Helson, H. (1964) ADAPTATION LEVEL THEORY. Harper & Row.

Ladefoged, P. & Broadbent, D. E. (1957) "Information conveyed by vowels," JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, 29:98.

Massaro, D. (1975) UNDERSTANDING LANGUAGE, Academic Press.

Miller, R. L. (1953) "Audiology tests with synthetic vowels," JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, 25:117.

Morse, P., Kass, J.E., & Turkienicz, R. (1976) "Selective adaptation of vowels," PERCEPTION & PSYCHOPHYSICS, 19:137-143.

Parducci, A. (1975) "A range-frequency analysis," in E. C. Carterette & M. P. Friedman (eds.), HANDBOOK OF PERCEPTION, Vol. II.

Restle, F. (1978) "Assimilation predicted by adaptation-level theory with variable weights," in N. J. Castellan & F. Restle (eds.), COGNITIVE THEORY, Vol. III.

Sawusch, J. R. & Nusbaum, H. C. (1979) "Contextual effects in vowel perception I: Anchor-induced contrast effects," PERCEPTION & PSYCHOPHYSICS, 25:292-302.

Sawusch, J. R. & Pisoni, D. B. (1976) "Response organization and selective adaptation to speech sounds," PERCEPTION & PSYCHOPHYSICS, 20:413-418.

Simon, H. J. & Studdert-Kennedy, M. (1978) "Selective anchoring and adaptation of phonetic and nonphonetic continua," JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, 64:1338-1357.

Slawson, A. W. (1968) "Vowel quality and musical timbre as functions of spectrum envelope and fundamental frequency," JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, 43:87-101.

Thompson, C. L. & Hollien, H. (1970) "Some contextual effects on the perception of synthetic vowels," LANGUAGE & SPEECH 13:1-13.