Testing the speech unit hypothesis with the primed matching task: Phoneme categories are perceptually basic

STEF DECOENE

Katholieke Universiteit Leuven, Leuven, Belgium

The basic speech unit (phoneme or syllable) problem was investigated with the primed matching task. In primed matching, subjects have to decide whether the elements of stimulus pairs are the same or different. The prime should facilitate matching in as far as its representation is similar to the stimuli to be matched. If stimulus representations generate graded structure, with stimulus instances being more or less prototypical for the category, priming should interact with prototypicality because prototypical instances are more similar to the activated category than are low-prototypical instances. Rosch (1975a, 1975b) showed that, by varying the matching criterion (matching for physical identity or for belonging to the same category), the specific patterns of the priming $\times$ prototypicality interaction could differentiate perceptually based from abstract categories. By testing this pattern for phoneme and syllable categories, the abstraction level of these categories can be studied. After finding reliable prototypicality effects for both phoneme and syllable categories (Experiments 1 and 2), primed phoneme matching (Experiments 3 and 4) and primed syllable matching (Experiments 5 and 6) were used under both physical identity instructions and same-category instructions. The results make clear that phoneme categories are represented on the basis of perceptual information, whereas syllable representations are more abstract. The phoneme category can thus be identified as the basic speech unit. Implications for phoneme and syllable representation are discussed.

What is the object to which speech perception processes respond? Which are the fundamental parts on which these processes work? These questions are of extreme importance in speech science. Knowing what the basic processing unit in speech perception is has important implications for the study of speech segmentation and invariance. Moreover, the development of a coherent, computational speech perception theory presupposes an empirically supported decision concerning the sequencing of representations resulting from speech perception processes.

Several such units have been proposed, but most research has focused on phonemes and syllables. However, until now, no clear consensus about the speech unit hypothesis has been found. I will argue that this is due largely to the use of an unsuitable experimental procedure: the monitoring task. In the research reported here, the primed matching task will be proposed and used to investigate the speech unit hypothesis, because it allows for a direct test of the status of the phoneme and the syllable as potential speech unit candidates.

The research reported here is part of a doctoral dissertation project (supervisor: G. Govaerts). Thanks go to G. van der Steene for making available PC/ATs, N. Bovens, R. Delabastida, and M. Lenaerts for technical assistance, and J. Kingston, J. Baele, and an anonymous reviewer for useful remarks. I am grateful to Tine Daeseleire for lending me her voice and to Bart Defoor and Jo Baele for help in running Experiment 2. Correspondence should be addressed to S. Decoene, Centrum voor Mathematische Psychologie en Psychologische Methodologie, Katholieke Universiteit, Tienestraat 102, 3-000 Leuven, Belgium (e-mail: fpaag08@blekul11.bitnet).

THE SPEECH UNIT HYPOTHESIS

The impetus for the investigation of the speech unit hypothesis came from a simple experiment. Savin and Bever (1970) had subjects monitor a sequence of meaningless syllables for either a complete syllable or a phoneme (the syllable-initial consonant or the medial vowel). Phoneme monitoring was significantly slower than syllable monitoring. Savin and Bever concluded that phonemes are identified after syllables of which they are parts, because phonemes are not perceptual, but abstract, nonsensory entities.

However, Foss and Swinney (1973) showed that response times for monitoring two-syllable words were faster than for word-initial syllables, and that syllables were responded to faster than were word-initial phonemes. Instead of arguing that syllables are more abstract than words, they denied the feasibility of dividing perceptual units into "abstract" and "real" ones on the basis of the monitoring task. Making a distinction between perception and identification, Foss and Swinney argued that the monitoring task does not tap into a processing level corresponding to immediate perception: The observed ordering of response times reflects the order by which linguistic units are (consciously) identified rather than the order in which they are perceived.

Although McNeill and Lindig (1973) subsequently showed that monitoring differences of linguistic segments depended on the match or mismatch between target and

search list, the monitoring task was not abandoned as an inappropriate method for investigating the speech unit hypothesis. On the contrary, it was used again and again, and various variables or artifacts were found to influence response times—for example, the identifiability of phonemes (Healy & Cutting, 1976), listeners' expectancy of target vowel context as determined by the match or mismatch between target and stimulus context (Mills, 1980a, 1980b) and uncertainty of vowel contexts (Swinney & Prather, 1980), the interaction between segmental and suprasegmental cues during ongoing perception (Meltzer, Martin, Mills, Imhoff, & Zohar, 1976), meaningfulness of stimuli (Rubin, Turvey, & van Gelder, 1976), the syllabic relationship between target and stimulus word (Mehler, Dommergues, Frauenfelder, & Segui, 1981; Segui, 1983), methodological confusion between syllable and phoneme monitoring (Mehler, Segui, & Frauenfelder, 1981), and the presence of foils (Norris & Cutler, 1988). The entire set of possible conclusions has been put forward on the basis of these experiments: The phoneme is the speech unit (Norris & Cutler, 1988), the syllable is the speech unit (Mehler, Dommergues, et al., 1981; Mehler, Segui, & Frauenfelder, 1981), the syllable and the phoneme are equally basic in speech perception (Healy & Cutting, 1976), and the monitoring task is inappropriate for studying the speech unit hypothesis (Mills, 1980a, 1980b; Rubin et al., 1976; Swinney & Prather, 1980).

Wood and Day (1975) used a two-choice speeded-classification task in which subjects had to identify either the consonant or the vowel of a monosyllable in which the irrelevant segment was either varied or held constant. They assumed that parameter variation will not affect response times when it does not occur within the same minimal perceptual unit. They found that response times were faster when the unattended segment was held constant, so it was concluded that the syllable was the minimal perceptual speech unit. However, Shand (1976) subsequently showed that variation of the first unattended syllable of a bisyllabic stimulus also affected response times to the consonant or vowel of the second syllable. By consequence, he concluded that Wood and Day's (1975) assumption was wrong and that monitoring in a selective attention task is not designed for the investigation of the speech unit hypothesis.

It should be stressed that the preceding discussion does not imply that the monitoring task is useless. Recent research has made it clear that the monitoring task is a powerful method for studying, for example, lexical processes (e.g., Cutler, Mehler, Norris, & Segui, 1987; Foss & Blank, 1980; Frauenfelder & Segui, 1989; Frauenfelder, Segui, & Dijkstra, 1990). My point is more modest: The monitoring task is not sensitive to the potential ordering (from concrete to abstract) of representations available during speech perception and, thus, is inappropriate for investigating the speech unit hypothesis.

Recently, the interpretation of this body of data has become even more complicated. In an interesting series of

monitoring studies, Cutler and colleagues (Cutler, Mehler, Norris, & Segui, 1983, 1986, 1989) obtained language-specific syllable monitoring effects. French, which has a simple syllable structure, generated clear syllabic segmentation, whereas English, which has unclear syllabic boundaries due to ambisyllabic segments, showed no syllabic segmentation.¹ But how can language-specific segmentation strategies be related to any systematicity in speech production or to any uniform representational format in speech perception? Cutler and Norris (1985) argued that it is not the classification of speech objects, but the segmentation process itself that should be made the central topic: Speech perception processes are designed to "utilize any information which can be extracted (either directly or indirectly) from the signal to determine where lexical boundaries lie" (p. 698). That is, the results of Cutler et al. (1983, 1986, 1989) make sense only if attention is shifted from speech perception to lexical representation and, more specifically, to the unit of lexical access. One can accept this solution on practical grounds. But two elements should be made explicit. First, the core of their solution is the definition of the term *segmentation* as lexical boundary detection, which is an extension of its original meaning in speech perception research: detection of segmental boundaries in the acoustic signal, that is, detection of the speech objects. Second, it supports the methodological moral already made by Foss and Swinney (1973): The monitoring task is not designed to investigate the nature of speech object representation, and consequently, it cannot be used to test the abstraction level of a possible speech perception object representation.

Although the monitoring task has been the predominantly used method for studying the speech unit hypothesis, some data do exist that are not based on it. Starting from the concept of a preperceptual auditory image and the time it is available for processing, Massaro (1972, 1974) used a recognition-masking task, finding that the processing of speech stimuli occurs within 200–250 msec after stimulus presentation—a duration of roughly syllable length. However, because this represents the upper boundary of speech unit length, smaller units, such as demisyllables (proposed by Samuel, 1989, on the basis of a selective adaptation study) or phonemes, would also fit in.

As a result of these diverging data, a certain amount of relativism concerning the speech unit hypothesis has been growing (e.g., Jusczyk, 1986; Pisoni & Luce, 1987). According to this view, there is not one unit but many, depending on the level of processing during speech perception. Such relativism erroneously assumes that, since there are many levels, the search for the basic speech unit is misguided. This assumption is wrong because the speech unit hypothesis does not ask whether, for example, phoneme categories are psychologically real—that is, present at some level of processing. There is a large amount of evidence in favor of segmental representations in language processing (e.g., see Pisoni & Luce, 1987, for a review). Also, the speech unit hypothesis is not concerned with the question of how the lexicon is accessed.

It is not necessarily the case that, for example, because the syllable is the basic speech perception unit, lexical contact representations also must be syllables. In my view, the assumption that the speech unit and the lexical contact representation are the same is an unfortunate one. The speech unit problem is a topic for speech perception research, not for the study of word recognition and on-line language comprehension.

Essentially, the speech unit hypothesis asks which representation among all the potential representations that can be (and are effectively) used during speech perception (phoneme category, syllable category, word, etc.) describes best the basic "building blocks" of speech. The category representation of these basic building blocks will be directly perceptually based, because it is directly determined by the acoustic information in the signal. And, since the abstraction level of mental categories is assumed to be closely related to the amount of direct perceptual determination of their representational content, the mental representation corresponding to the basic speech segment should be the most concrete mental category.

Given the fundamental importance of the speech unit hypothesis for speech perception theories, one should concentrate on the search for an experimental procedure allowing a more direct investigation of this problem, instead of shifting attention to lexical processing. Such a procedure should fit the logic of the speech unit hypothesis, which centers around the notion of the abstraction level of representations. In the following section, I will discuss the primed matching task as a procedure that fits this logic.

Before I discuss the primed matching task, a small note on terminology may be in order. In the subsequent discussion, I will use the terms *phoneme category* and *syllable category*, instead of *phonetic segment* and *unit of roughly syllabic length*. In this way, I want to indicate that the focus of this research is on how these two entities are mentally represented, and not on issues of whether, for example, speech perception entails phonetic processing, central auditory processing, or both. That is, the topic of the research reported here is speech representation, not speech processing. While processing issues are important, they do not stand in the foreground.

THE PRIMED MATCHING TASK: A METHOD FOR DIFFERENTIATING CATEGORIES AT VARIOUS LEVELS OF ABSTRACTION

Mental categories can be placed on a continuum from concrete to abstract, from entirely perceptually defined to entirely conceptual (Medin & Barsalou, 1987; Rosch, 1975a, 1975b): The color blue is a perceptually based category, whereas the concept "freedom" has a purely abstract characterization. Although perceptual and conceptual categories are highly similar (see Medin & Barsalou, 1987, for an excellent discussion), there is an interesting

finding using the primed matching task, as discussed in Rosch (1975a, 1975b), on which perceptual and more abstract categories seem to be differentiable.

The Primed Matching Task

In the primed matching task (Beller, 1971), subjects are asked to decide as rapidly as possible whether or not elements of a stimulus pair are the same. The priming stimulus preceding the stimulus pair should facilitate the matching process insofar as the mental category representation it activates is similar to the stimuli to be matched. If stimuli are instances of a category that generates a prototypicality effect, then the prime will activate a category representation that is more similar to a good example than to a bad example of the category. Thus, the prime will facilitate judging similarity more in matching good-example pairs than in matching poor-example pairs. Primed matching will thus result in a priming $\times$ prototypicality level interaction (henceforth, a $P \times P$ interaction).

Moreover, subjects can be induced to match stimuli using different criteria by instructing them to judge stimuli of a pair as the same if they are physically identical or as the same if they belong to the same category. The latter criterion allows subjects to use any information available; the former restricts subjects to using only direct, low-level perceptual information.

Rosch (1975a, 1975b) showed that application of the priming logic to this variation in matching instructions allowed for the differentiation of categories that are based on perceptual information from categories that are described by abstract, nonperceptual information, given that they have graded structure. If subjects have to match stimuli under physical identity instructions, the prime will be effective only if the activated category representation contains low-level perceptual information, thus enabling faster matching by the use of this low-level perceptual information (resulting in a $P \times P$ interaction). If matching has to be done under same-category instructions, different predictions concerning a $P \times P$ interaction are possible, depending on the representational content of the category representation (e.g., entirely perceptually determined, entirely based on abstract information, or based on a mixture of perceptual and more abstract information). For example, if the prime activates only perceptual information, then same-category pairs will not show a $P \times P$ interaction when same-category instructions are used, because they do not match with respect to the activated low-level perceptual information.

The research of Rosch (1975a, 1975b) showed the strength of the primed matching task. Rosch (1975b, Experiment 1) found a $P \times P$ interaction for matching physically identical instances of color categories under physical identity instruction. In a complementary experiment (Rosch, 1975b, Experiment 3), matching under same-category instructions returned a significant $P \times P$ interaction for physically identical color pairs but no $P \times P$ interaction for same-category pairs. Within the context of the primed

matching logic, these results strongly supported a graded structure representation for color categories with representation primarily based on perceptual information.

An entirely different pattern was found when the representations generated by superordinate semantic category names were studied (e.g., "chair"). Under physical identity instructions (Rosch, 1975a, Experiment 5), no $P \times P$ interaction was found for physically identical pairs. Under same-category instructions (Rosch, 1975a, Experiment 2), a significant $P \times P$ interaction for physically identical pairs was observed, but no such interaction was observed for same-category pairs. From this pattern of results, Rosch (1975a) inferred that the mental representation of semantic categories is based on a more abstract (semantic) code than is, for example, that of color categories.

The primed matching task, then, promises to be a method better suited to investigate the abstraction level of representations in speech perception than the phoneme/syllable monitoring task. Rosch (1975a, pp. 226-227) explicitly pointed out that the primed matching task made possible "a detailed investigation of the nature of the mental representations generated by category terms at any level of abstraction." The task has the advantage that it does not depend on the comparison of average response times to phoneme or syllable categories (as is the case with phoneme/syllable monitoring), but instead compares them on a specific pattern over experimental manipulations. Moreover, the task focuses directly on the representation of phoneme and syllable categories and thus avoids the possible confounding of representation versus on-line perceptual processing of speech—the primed matching task does not test which unit is made (consciously) available first, but it does test what the content of a category representation is (perceptually determined or not). No confusion between perceptual processes (which are consciously inaccessible) and conscious identification (Foss & Swinney, 1973) is possible.

Phoneme and Syllable Prototypicality Effects

As described above, the crucial prerequisite when using the primed matching task is that phonemes and syllables should be categories generating graded structure.

As far as phoneme categories are concerned, empirical evidence supports the idea that they have a representational structure that generates prototypicality effects. Miller, Connine, Schermer, and Kluender (1983) used a selective adaptation procedure to study the possible auditory basis of internal phonetic category structure. Syllable series for three phonetic features—voicing, place of articulation, and manner of articulation—were synthesized. They found adaptation functions approximating an inverted U. Miller et al. hypothesized that the adaptation peaks corresponded to the location of the phonetic prototype on the stimulus continuum. However, they did not have a measure of example goodness for the category instances that would have allowed them to test the hypothesis directly. Samuel (1982) made such an external measure by letting subjects perform both a response-timed

identification task and an overt prototype task on a synthesized subset of the /ga/-/ka/ continuum. He found that prototype adaptors produced more adaptation than did other adaptors. Miller and Volaitis (1989) observed a clear relation between ratings of how good a /p/ the consonants of two synthesized /bi/-/pi/ series with progressively larger VOT values were and the effect of syllable duration between the two series on an identification task—as expected, an inverted U-shaped function was found for each series, while the increase of syllable duration both shifted the location of the good exemplars and widened their range.

The best evidence to date in favor of prototypicality effects for phoneme categories comes from research by Kuhl and colleagues. Using a head-turn procedure, Grieser and Kuhl (1989) trained infants of about 6 months of age with either a synthesized good or a poor example of the vowel /i/ (as defined by adults). They subsequently obtained better generalization to novel instances of the vowel for infants trained with the good example (see also Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992). Kuhl (1991) had subjects (both adults and infants) decide when a referent speech sound (either a synthesized good or a poor example of the vowel /i/) presented continuously once per second changed to a comparison speech sound. A *perceptual magnet effect* for phoneme category prototypes on stimuli nearby in perceptual space, was demonstrated: Prototypical speech sounds assimilated other instances, making the latter perceptually more similar to the former.

In the present research, natural phoneme categories occurring word-initially were used to test for the presence of a prototypicality effect. Since speech perception is cued by a multitude of information sources, it would be more interesting to find graded structure for natural segments in which natural variation is the dominant factor. Indeed, it seems likely that an ecologically valid speech perception phenomenon, such as a prototypicality effect, will be based on exactly such variation (with prototypicality determined by, for example, phonetic context). Moreover, it has been shown that well-established effects in speech perception can disappear when more natural stimuli are used (e.g., see Burton, Baum, & Blumstein, 1989, for the influence of a more natural stimulus structure on lexical effects in phonetic categorization). As will be discussed later, the problems that could be created in turn by using natural stimuli can be dealt with by the primed matching task.

As far as syllables are concerned, evidence for graded category structure is sparse. Some models incorporate syllable prototypes by assumption—notably, the fuzzy logical model of speech perception (Massaro, 1987, 1989; Oden & Massaro, 1978). However, as far as I know, direct empirical support for the existence of syllable prototypes does not exist.

In summary, there is good evidence that phonemes are categories whose representations generate graded structure. However, the prototype view would be supported more strongly, and the speech unit problem could be investigated more decisively within a categorization perspec-

tive, if reliable prototypicality judgments could be tapped with natural phoneme categories. This possibility was evaluated in Experiment 1. In Experiment 2, possible prototypicality effects for syllable categories were studied.

A DIRECT TEST OF THE SPEECH UNIT HYPOTHESIS WITH THE PRIMED MATCHING TASK

The following experiments use the primed matching task under both physical identity and same-category instructions in order to provide a direct investigation of the speech unit hypothesis. First, the reliability of prototypicality ratings for natural phoneme and syllable categories were assessed (Experiments 1 and 2). These prototypicality judgments were the basis for four experiments using the primed matching task. In Experiments 3A and 3B, subjects performed a primed phoneme matching task under physical identity instructions and under same-category instructions. In Experiments 4A and 4B, a primed syllable matching task was performed under physical identity and same-category instructions.

There are two crucial questions. First, can a significant $P \times P$ interaction be obtained, demonstrating that good-example and poor-example pairs are affected differently by priming, with either primed good-example pairs responded to faster than unprimed, primed poor-example pairs responded to slower than unprimed, or both? If the answer to this question is negative for phoneme and/or syllable categories, it should be concluded that the goodness-of-example ratings obtained in Experiment 1 and/or Experiment 2 are merely artificial, and do not correspond to any "real" differences between good and poor examples of a category. By consequence, it would be impossible to conclude anything at all about the speech unit hypothesis.

If priming has a different effect on good-example and poor-example pairs, the second question is: Under which instructions (physical identity and/or same-category), with which stimulus pairs (physically identical pairs and/or same-category pairs), and with which categories (phonemes and/or syllables) can a $P \times P$ interaction be observed? An analysis of the specific distribution of these $P \times P$ interactions should make clear whether one, both, or none of the category types is perceptually based. For example, if syllable categories are perceptually based, one should expect a $P \times P$ interaction for physically identical pairs under both physical identity instructions and same-category instructions, but no such interaction for same-category pairs under same-category instructions.

Two additional, important points should be made. The fact that the experiments of Rosch described above point at a way to differentiate mental representations as far as their level of abstraction is concerned should not blur the differences with the present study. The stimuli used by Rosch (1975a, 1975b) were perceptual categories, on the one hand, and semantic categories, on the other hand. In the present study, the stimuli of interest, phoneme versus

syllable categories, are both presented as spoken speech stimuli. As such, they can both be described as perceptual. Thus, the investigation here does not ask whether phoneme or syllable categories are perceptual categories (that would be a truism), but whether and how directly their mental representation is based on purely perceptual information.

Second, searching for graded structure has no processing implications per se (see Lakoff, 1986, and Rosch, 1978, for discussion of this point). Finding graded category structure is compatible with various theories of categorization, such as representation by prototypes or by exemplars. Reliable and valid graded structure (e.g., for speech categories) does not uniquely constrain the kind of categorization model that should be defended. The research reported here has nothing to say about these issues. The crucial question is whether or not reliable and valid graded structure can be found.

EXPERIMENT 1 Prototypicality Ratings of Phoneme Categories

Method

Stimuli. Seven Dutch phoneme categories were chosen: three consonants (/k/, /r/, and /s/) and four vowels (/ɛ/, /a:/, /a/, and /e:/). The categories and the words are part of the stimulus set currently in use at the Katholieke Universiteit Leuven laboratory of auditory perception and speech communication. This stimulus set had originally been constructed to study the perceptual effects of allophonic distribution. Consequently, each set incorporated all possible allophones for a given phoneme. All stimulus words were low frequency (Uit den Boogaert, 1975), with a median frequency of 0; the crucial phoneme categories occurred in word-initial position. There were 240 words in the /k/ set, 215 words in the /r/ set, 98 words in the /s/ set, 67 words in the /e/ set, 40 words in the /a:/ set, 124 words in the /a/ set, and 90 words in the /ɛ/ set.

The stimulus words were recorded twice during 10 sessions over a 1-week period in a sound-proof room, with a Stellavox SP7 recorder and Sennheiser MD441N microphone. A female speaker pronounced the words with flat intonation. When the entire set was recorded, the speaker listened to the two versions and selected the best. Whenever the speaker felt unsure about the quality of the recorded material, the words were recorded again. Stimulus tapes were then constructed by randomizing words belonging to a category. An interval of about 5 sec separated each stimulus word in order to give the subjects enough time to give their goodness-of-example ratings.

Subjects. The subjects were first-year psychology and educational sciences students. They participated in partial fulfillment of course credits. None reported hearing deficiencies. A group of 34 subjects rated the /s/ set, the /e:/ set, the /k/ set, and the /a:/ set. Another group of 43 subjects rated the /a/ set, the /ɛ/ set, and the /r/ set. One subject did not complete his booklet with ratings for the /a/ set.

Procedure. Each group of subjects listened to the stimulus tapes in a room with excellent acoustics. They received a booklet containing rating scales with 10 intervals in graphical form combined with numeric symbols and anchored at the endpoints (Nunnally, 1970). The subjects first received elaborate written instructions. They were told that prototypicality could be used to judge speech segments such as phoneme categories, and they were instructed to listen carefully to all stimulus words and to indicate how good a member the relevant segment of each word was for the phoneme category under consideration. When the subjects still felt unsure after reading the instructions, the experimenter gave an example,

using a segment not included in the experimental set (/l/). Pauses of about 5 min were given between each category set. For each group, the experiment lasted about 1 h.

Results and Discussion

The subjects' ratings for the individual phoneme categories used all steps of the scale, with ratings varying from 1 to 10, and overall median ratings of 7 for /k/, /e:/, /a:/, and /s/, 6 for /ɛ/ and /a/, and 5 for /r/.

Two indices of interobserver agreement were computed. First, Stine's (1989) omicron coefficient (corrected for chance) of the median ratings obtained between split halves of the sample of subjects, divided at random, was calculated. Stine's omicron is an index of relational agreement between observers; it measures the degree to which judgments represent admissible transformations of the ordinal relations between the stimuli. High ordinal relational agreement means that the ordinal relations between stimuli are preserved over the observers. Also, the intraclass correlation ICC(2, *k*) (Schout & Fleiss, 1979) was calculated. The intraclass correlation measures the correlation between the median ratings of *k* observers, treating observers as a random effect. Both the intraclass correlation and omicron were significant for all categories (see Table 1). The intraclass correlations varied from .663 for /s/ to .862 for /k/. The values of omicron ranged from .46 for /s/ to .837 for /a:/.

It could be objected that although the ratings found in Experiment 1 were reliable, nothing is known about their actual basis. Are they real, or just artificial or speech irrelevant? And even if they are real, how exactly were they generated?

It goes without saying that an investigation of the basis of prototypicality effects for word-initial natural phoneme categories is interesting in its own right. However, it should be stressed that an answer to this question does not affect the interpretation of potentially significant $P \times P$ interaction effects in the primed matching task, because the design of the primed matching task allows one to detect any spurious effects of goodness level. (1) If the reliable prototypicality effects for phoneme categories, for example, are mere response strategies of subjects, neither a significant $P \times P$ interaction nor a significant main effect of goodness level can be expected. (2) If the reliable

prototypicality ratings are determined by systematic but speech-irrelevant information (such as loudness differences between phoneme category instances, or the confounding of the phoneme category ratings with, for example, lexical factors), the most one can expect to find in the primed matching task is a significant main effect of goodness level, but no significant $P \times P$ interaction. Cross-modal priming with a category name activates information related to the category representation. So, if prototypicality effects found for phoneme categories are not based on information used in the description of the category representation, priming will not differentially affect good and poor examples of a category. A significant $P \times P$ interaction can be found if and only if there is such a differential match between the primed category representation and good-example versus poor-example category instances. This implies that the empirical distinction between good-example and poor-example category instances is not only reliable but also valid. In strictly operational terms, by looking for an interaction specified a priori between priming and prototypicality, interpretation of the prototypicality effect is much more constrained (via the priming logic) than when only a main effect of prototypicality would be predicted.

In summary, since it can be safely concluded from Experiment 1 that reliable goodness-of-example judgments can be obtained for natural phonemic categories with a direct rating procedure, Experiments 3A and 3B were run to assess the processing implications of these goodness-of-example ratings and to test the speech unit hypothesis (if, at least, these goodness-of-example judgments are valid).

EXPERIMENT 2

Prototypicality Ratings of Syllable Categories

Method

Stimuli. Eleven syllable categories, which are part of the stimulus set used at the laboratory, were used. These syllables had been chosen so as to include members of each of the four simple syllables acceptable in Dutch (Clements & Keyser, 1983; Collins & Mees, 1981; van der Hulst, 1984). These four elementary syllable types were a vowel (V) syllable, a consonant-vowel (CV) syllable, a vowel-consonant (VC) syllable, and a consonant-vowel-consonant (CVC) syllable. The stimulus set included /o:/ (*n* = 53) and /a/ (*n* = 161) as V syllables, /be/ (*n* = 131), /na:/ (*n* = 43), and /pa/ (*n* = 89) as CV syllables, /op/ (*n* = 90), /as/ (*n* = 67), and /uit/ (*n* = 105) as VC syllables, and /kop/ (*n* = 77), /vɛt/ (*n* = 58), and /vak/ (*n* = 83) as CVC syllables. Only two V syllables were available, because it was difficult to find categories of this syllable type with enough instances. All syllable category instances occurred in word-initial position of two-to-four-syllable low-frequency words (median frequency = 0; Uit den Boogaert, 1975). The recording procedure was identical to that of Experiment 1. The same female speaker was used to pronounce the syllables.

Subjects. The subjects were first-year psychology and educational sciences students. They participated in partial fulfillment of course credits. None reported hearing deficiencies. A group of 45 subjects rated the /na:/ set, the /as/ set, the /kop/ set, the /vɛt/ set, the /o:/ set, the /uit/ set, and the /op/ set. Another group of 25 subjects rated the /be/ set, the /pa/ set, the /vak/ set, and the /a/ set. Each session lasted about 1 h.

Table 1
Interobserver Agreement for Goodness-of-Example
Ratings of Phoneme Categories

Phoneme	<i>S</i>	<i>n</i>	ICC (2, <i>k</i>)	<i>F</i>	Omicron	<i>t</i>
/a:/	40	34	.771	5.5	.837	11.7
/a/	124	42	.756	4.6	.585	5.7
/e:/	67	34	.803	6.6	.741	8.2
/ɛ/	90	43	.822	6.4	.730	9
/s/	98	34	.663	3.7	.460	3.6*
/k/	240	34	.862	8.4	.779	9.4
/r/	215	43	.788	5.3	.696	8.1

Note—*S* = number of stimuli. *n* = number of subjects. All values of ICC (2, *k*) and omicron are significant, $p < .0005$, unless indicated otherwise. * $p < .005$.

Procedure. The procedure was the same as in Experiment 1. If necessary, the experimenter provided an additional example (/bal/).

Results and Discussion

The subjects' ratings for the individual syllable categories used all steps of the scale, with ratings varying from 1 to 10, and overall median ratings of 7 for /o:/, /be/, /na:/, /uit/, /op/, /as/, and /vet/, 6.5 for /pa/, /vak/, and /a/, and 6 for /kop/.

Both the intraclass correlation ICC(2,*k*) and omicron were computed. Both indices obtained significance (see Table 2). The intraclass correlations varied from 0.395 for /a/ to 0.893 for /o:/. Omicron varied from 0.143 for /a/ to 0.868 for /op/.

It is clear from Tables 1 and 2 that the values of the interobserver indices were comparable to those of phoneme categories. These results indicate that the subjects were able to reliably judge the goodness of example for syllable categories with a direct prototypicality rating task. There was one deviation from the overall pattern. The omicron value for the /a/ set was, though significant, too low. It is not clear why this syllable category behaved differently from the others. It is possible that the group of subjects who rated the /a/ set had difficulty perceiving the stimuli as syllables. However, the good results for the /o/ set indicate that this difficulty has to be attributed to this particular group of subjects. A small replication was run, in which 32 subjects rated the stimuli of the /a/ set, under identical instructions and with the same procedure. Interobserver agreement was clearly better [ICC(2,32) = .480, $F(160,4960) = 2.24$, $p < .0005$; omicron = .320, $t(159) = 4.26$, $p < .0005$]. Hence, the /a/set was not dropped.

In sum, because these prototypicality ratings were reliable, they could be used to investigate the processing implications of the goodness-of-example ratings and to test the speech unit hypothesis (if, at least, these goodness-of-example judgments are valid) (Experiments 4A and 4B).

Table 2
Interobserver Agreement for Goodness-of-Example
Ratings of Syllable Categories

Syllable	<i>S</i>	<i>n</i>	ICC(2, <i>k</i>)	<i>F</i>	Omicron	<i>t</i>
/o:/	53	45	.893	12.7	.798	9.5
/a/	161	25	.395	1.9	.143	1.8*
/a/‡	161	32	.480	2.2	.320	4.3
na:/	43	45	.773	6.5	.414	8.2†
/pa/	89	25	.548	2.7	.507	5.6
/be/	131	25	.720	4.3	.481	6.2
/as/	67	45	.730	4.9	.629	6.5
/uit/	105	45	.705	4.3	.581	7.2
/op/	90	45	.798	7.4	.868	16.5
/kop/	77	45	.869	9.8	.838	13.3
/vak/	83	25	.763	5.0	.447	4.5
/vɛt/	58	45	.828	7.5	.674	6.8

Note—*S* = number of stimuli. *n* = number of subjects. All values of ICC(2,*k*) and omicron are significant, $p < .0005$, unless indicated otherwise. * $p < .05$. † $p < .005$. ‡Results of the replication with the /a/ set.

EXPERIMENT 3A Primed Phoneme Matching Under Physical Identity Instructions

The results of Experiment 1 form a reliable basis for using the primed matching task with phoneme categories. Experiment 3A used the primed masking task under physical identity instructions; Experiment 3B used the primed matching task under same-category instructions. The results of these two experiments will be discussed together.

Method

Subjects. Twelve second-year psychology students participated in the experiment in partial fulfillment of course credits. None reported hearing deficiencies.

Stimuli. The stimulus words used here were chosen from the words used in Experiment 1. For each phoneme category, stimulus words were rank ordered from high to low prototypical, and the highest and the lowest prototypical stimulus word was selected (resulting in 14 stimulus words). Pairs of stimulus words were then constructed according to the procedure described in Rosch (1975a, 1975b).

For each of the seven phoneme categories, a pair of physically identical prototype instances (e.g., [kaars-kaars]) and a pair of physically identical, poor-example instances (e.g., [kwezel-kwezel]) was constructed. Additionally, 14 different stimulus pairs were made: 7 pairs in which the prototypical member of one category was combined with the prototypical member of another category (e.g., [kaars-siroop]), and 7 pairs in which the nonprototypical member of one category was combined with the nonprototypical member of another category (e.g., [kwezel-elatiet]; see Appendix A for the complete stimulus set). All in all, there were 28 stimulus pairs.

Experimental setup. The stimulus pairs were randomized and recorded on one track of a two-track recorder (Philips AAC 4000 language trainer). Each stimulus pair occurred twice: once preceded by the orthographic presentation of the phoneme category name of the first category (the priming stimulus), and once preceded by an exclamation mark (giving a total of 56 stimulus pairs). Primed and unprimed trials were alternated (Beller, 1971; Rosch, 1975a, 1975b). Within each pair, the stimulus words were recorded with an interval of about 250 msec so that forward masking could be avoided (Scharf & Buus, 1986). Stimulus pairs were separated by an interval of 5 sec.

A short tone pulse was put on the other track of the tape about 2 sec before the second word of the pair. The exact time interval between pulse location and onset of the second stimulus word was controlled using the Philips speech adapter box and OM8210 speech-processing software, so that pulse location could be corrected if necessary. The distance between prime and the first stimulus word was 750–1,000 msec. This falls within the time interval in which the benefit or cost (in milliseconds) of priming remains constant when compared with 2 sec (Rosch, 1975b), and thus ensured that encoding of the first phoneme category in each pair would also be affected sufficiently by the prime.

The tone pulse was input to a computer (PC/AT), and it activated both the presentation of the prime on a monochrome computer display and the response-time measurement. The prime was presented for 500 msec. A Turbo Pascal program (Bovens & Brysbaert, 1990) was used for the response-time measurement.

Procedure. In this and the following experiments, the subjects participated in groups varying from 2 to 5. They were seated in a sound-attenuated room before a computer to which two key buttons were connected. They were instructed to press the *same* key

if the words in the stimulus pair began with physically identical segments and the *different* key for different word-initial segments; they were asked to do this as rapidly as possible without making errors. Half of the subjects used the dominant hand for the *same* key, and half used the dominant hand for the *different* key. The subjects listened to the stimulus tape over headphones (Bayer H 805). Four practice pairs (two identical and two different) were given in order to familiarize the subjects with the procedure and the judgment to be made. The experiment lasted about 20 min.

Results

Overall error rate was 3%. Within-subject, three-way analyses of variance (ANOVAs; randomized block factorial design; see Kirk, 1982; Winer, 1971), with level of goodness of example, priming, and phoneme type (vowel/consonant) as fixed effects, and subjects as a random effect, were performed on the response times for same and different pairs separately. The critical priming $\times$ goodness level interaction for physically identical pairs and for different pairs are depicted in Figure 1.

For pairs of physically identical stimuli, the ANOVA revealed a significant $P \times P$ interaction effect [$F(1,77) =$

7.73, $p < .01$; see Figure 1A]. Primed prototypical pairs were responded to faster (358 msec) than were unprimed prototypical pairs (426 msec). A planned t test² (Keppel, 1982; Kirk, 1982) was computed, and it showed that this difference was significant [$t(46) = 1.79$, $p < .05$]. The subjects also responded slower to primed poor-example pairs (417 msec) than to unprimed poor-example pairs (377 msec). However, the difference was not significant.

Two other effects attained significance. First, the phoneme type $\times$ priming interaction was significant [$F(1,77) = 15.74$, $p < .001$]. A posteriori Tukey tests ($p = .01$) within category types on the simple main effect were computed. Differences were not significant. Second, the main effect of phoneme type showed that vowels were responded to faster than were consonants [352 msec vs. 452 msec; $F(1,77) = 24.39$, $p < .001$].

For pairs of different stimuli, only the type $\times$ goodness level interaction was significant [$F(1,77) = 5.74$, $p < .05$]. A posteriori Tukey tests ($p = .01$) of the simple main effect revealed no significant differences within category types.

EXPERIMENT 3B Primed Phoneme Matching Under Same-Name Instructions

Method

Subjects. Thirteen second-year psychology students participated in partial fulfillment of course credits. None reported hearing deficiencies.

Stimuli. The stimulus pairs used in Experiment 3A were used again in Experiment 3B. In addition, for each category, a prototypical same-name pair (e.g., [kaars-kamer]) and a poor-example same-name pair (e.g., [kwezel-koegang]) were constructed (Rosch, 1975b). These stimulus words were the next best and poorest examples in the prototypicality ranking.

The stimulus set consisted of 42 stimulus pairs: a pair of physically identical prototype instances for each of the seven categories, a pair of identical poor-example instances for each category, a pair of prototypical same-name instances for each category, a pair of poor-example same-name instances for each category, 7 pairs of prototypical different instances, and 7 pairs of poor-example different instances (Appendix A).

The stimulus tape was constructed as described for Experiment 3A, consisting of 84 stimulus pairs.

Procedure. The experimental procedure was identical to that of Experiment 3A. The only exception was the instruction given: The subjects were asked to press the *same* key if word-initial segments of a word pair belonged to the same category and to press the *different* key if word-initial segments of a word pair were different. The experiment lasted about 30 min.

Results

Overall error rate was 4%. Within-subject, three-way ANOVAs (randomized block factorial design) on the response times, with goodness level, priming, and phoneme type (vowel/consonant) as fixed effects, and subjects as a random effect, were performed separately on the physically identical, same-name, and different pairs. The critical priming $\times$ goodness level interaction for physically identical, same-name, and different pairs is depicted in Figure 2.

Figure 1. Mean response times in milliseconds for the priming $\times$ goodness level interaction effect for primed phoneme matching under physical identity instructions: (A) physically identical stimulus pairs, (B) different stimulus pairs. Full line = primed pairs; broken line = unprimed pairs.

Figure 2. Mean response times in milliseconds for the priming $\times$ goodness level interaction effect for primed phoneme matching under same-category instructions: (A) physically identical stimulus pairs, (B) same-category stimulus pairs, (C) different stimulus pairs. Full line = primed pairs; broken line = unprimed pairs.

The ANOVA for the physically identical stimulus pairs showed a significant $P \times P$ interaction effect [$F(1,84) = 10.5, p < .005$; see Figure 2A]. Primed prototypical pairs of physically identical stimuli were responded to faster (345 msec) than were unprimed pairs (425 msec). A planned t test showed this difference to be significant [$t(50) = 1.84, p < .05$]. The difference between primed versus unprimed poor-example pairs was not significant (410 msec vs. 399 msec, respectively).

The priming $\times$ type interaction was significant [$F(1,84) = 4.3, p < .05$]. A posteriori Tukey tests ($p = .01$) on the simple main effect showed a significant difference between primed and unprimed vowels (353 msec vs. 420 msec, respectively). The main effect of priming was also significant [$F(1,84) = 4.02, p < .05$]. Primed pairs were responded to faster than were unprimed pairs (377 msec vs. 412 msec).

Same-category pairs behaved differently. Here, the $P \times P$ interaction was not significant (Figure 2B). The main effect of goodness level was significant [$F(1,84) = 9.61, p < .005$]. Good-example pairs were responded to faster than were poor-example pairs (493 msec vs. 560 msec). The phoneme type $\times$ priming interaction was also significant [$F(1,84) = 26.6, p < .001$]. Primed consonant pairs were responded to faster than were unprimed consonant pairs (482 msec vs. 587 msec), whereas primed vowel pairs were responded to slower than were unprimed vowel pairs (573 msec vs. 468 msec). Both differences were significant when tested with a posteriori Tukey tests ($p = .01$) of the simple main effect.

For pairs of different stimuli, the only significant effect was the main effect of goodness level. Pairs of different good-example stimuli were responded to faster than were pairs of different poor-example stimuli [531 msec vs. 599 msec; $F(1,84) = 8.97, p < .005$].

Discussion of Experiments 3A and 3B

The crucial finding of Experiments 3A and 3B is the presence of a significant $P \times P$ interaction, which allows for two important conclusions concerning phoneme categories.

First, priming occurs when processing of a stimulus is facilitated because of the (at least partial) information overlap between a previously activated representation (through presentation of the priming stimulus) and the representation of the primed stimulus. Consequently, the finding that good examples of phoneme categories are primed, whereas poor examples are not, supports the idea that phoneme categories exhibit graded structure. Thus, prototypicality ratings of natural phoneme categories (Experiment 1) corroborate previously found prototypicality effects (Grieser & Kuhl, 1989; Kuhl, 1991; Miller et al., 1983; Miller & Volaitis, 1989; Samuel, 1982).

Equally important are the implications for the speech unit hypothesis. In the primed matching task, the subjects were asked to judge as rapidly as possible whether two stimuli were identical, where identity was defined as physical identity in Experiment 3A and defined as belonging to the same category in Experiment 3B. Physical identity instructions force the subjects to compare the perceptual characteristics of the stimuli in a pair, while same-category instructions imply that more abstract information can be used by the subjects in comparing the stimuli. However, irrespective of instructions, only responses to physically identical pairs showed a significant $P \times P$ interaction. That is, even when subjects are allowed to use both perceptual and more abstract characteristics of phoneme category representations, priming interacts with prototypicality

level only for physically identical stimulus pairs. Only one conclusion seems possible: The content of phoneme category representations consists of concrete, sensory information only. On the basis of the distribution of the $P \times P$ interaction over stimulus pairs and instructions, there can be little doubt that phoneme categories are not purely abstract categories (as Savin & Bever, 1970, suggested) but that they are represented on the basis of perceptual information.

Finally, some other effects attained significance. The significant main effects of phoneme type in Experiment 3A and goodness level in Experiment 3B are of little importance since they do not affect the interpretation of the priming $\times$ goodness level interaction. The only results that need comment are the phoneme type $\times$ priming interactions, which attained significance for (1) physically identical pairs under physical identity instructions and (2) physically identical pairs and same-category pairs under same-category instructions. These results present a bit of a puzzle since there is no a priori reason to expect this interaction. However, since the three-way interaction of type $\times$ priming $\times$ goodness level was not significant, the $P \times P$ interaction can be interpreted in its own right³ (Keppel, 1982; Kirk, 1982; Winer, 1971).

In summary, Experiments 3A and 3B show that the basis of phoneme representations is perceptual and that the reliable prototypicality ratings of Experiment 1 refer to real, speech-characteristic differences between instances of phoneme categories, and not just to random or speech-irrelevant responses.

EXPERIMENT 4A Primed Syllable Matching Under Physical Identity Instructions

Using the reliable syllable prototypicality judgments obtained in Experiment 2, Experiments 4A and 4B used a primed syllable matching task under physical identity and same-category instructions to evaluate the representational status of syllable categories. The results of Experiments 4A and 4B will be discussed together.

Method

Subjects. Eleven second-year psychology students participated in partial fulfillment of course credits. None reported hearing deficiencies.

Stimuli. The stimulus words were chosen from those used in Experiment 2. For each syllable category, stimulus words were rank ordered from high to low prototypical, and the highest and the lowest stimulus was selected. The syllables were all well segmented. The stimulus set consisted of 44 pairs of stimulus words with the crucial syllable in word-initial position, selected following the same procedure as in Experiment 3A: a pair of physically identical good-example instances and a pair of identical poor-example instances for each of the 11 categories, 11 pairs of different good-example instances, and 11 pairs of different poor-example instances (Appendix B).

The stimulus tape was constructed as described in Experiment 3A, consisting of 88 stimulus pairs.

Procedure. The procedure was equivalent to that in Experiment 3A, the only difference being the instructions, which were

adapted to the use of syllable categories. Six subjects used the dominant hand for the *same* key; 5 subjects used the dominant hand for the *different* key. The experiment lasted about 30 min.

Results

Overall error rate was 3%. Within-subject, three-way ANOVAs (randomized block factorial design) on the response times, with goodness level, priming, and syllable type as fixed effects, and subjects as a random effect, were computed. Analyses for physically identical pairs and different pairs were performed separately. The critical $P \times P$ interaction for physically identical and for different pairs is depicted in Figure 3.

For physically identical pairs, no significant $P \times P$ interaction was found (see Figure 3A). A significant effect was found for the main effect of syllable type [$F(3,150) = 8, p < .001$]. V syllables were responded to faster than were CV and VC syllables, which differed only marginally, whereas responses to CVC syllables were slowest ($459 \text{ msec} < 501 \text{ msec} \approx 509 \text{ msec} < 545 \text{ msec}$). The type $\times$ goodness level interaction was also significant [$F(3,150) = 3.87, p < .05$]. However, a posteriori

Figure 3. Mean response times in milliseconds for the priming $\times$ goodness level interaction effect for primed syllable matching under physical identity instructions: (A) physically identical stimulus pairs, (B) different stimulus pairs. Full line = primed pairs; broken line = unprimed pairs.

Tukey tests ($p = .01$) of the simple main effect showed no significant differences within category types.

The ANOVA on different pairs revealed a significant syllable type $\times$ goodness level interaction [$F(3,150) = 11.63, p < .001$]. A posteriori Tukey tests ($p = .01$) of the simple main effect showed a significant difference within the V-syllable type conditions (545 msec for good-example pairs vs. 364 msec for poor-example pairs). The main effect of goodness level was also significant [$F(1,150) = 7.62, p < .01$]. Different good-example pairs were responded to slower than were different poor-example pairs (485 msec vs. 443 msec).

EXPERIMENT 4B Primed Syllable Matching Under Same-Name Instruction

Method

Subjects. Ten second-year psychology students participated in partial fulfillment of course credits. None reported hearing deficiencies.

Stimuli. The same stimulus pairs as in Experiment 4A were used, with additionally one prototypical same-category pair and one poor-example same-category stimulus pair for each of the 11 categories (Appendix B). These stimulus words were the next best and worst examples in the prototypicality ranking. Since each pair was presented twice—once primed, once unprimed—the stimulus set totaled 132 stimulus pairs.

Procedure. The experimental procedure was the same as in the previous experiments, the only difference being the instructions. The subjects were asked to press, as fast as possible, the *same* button if stimuli belonged to the same syllable category and to press the *different* button if stimuli belonged to different syllable categories. Five subjects used the dominant hand for the *same* key; 5 subjects used the dominant hand for the *different* key. The experiment lasted about 40 min.

Results

Overall error rate was 4%. Within-subject, three-way ANOVAs (randomized block factorial design) on the response times, with goodness level, priming, and syllable type as fixed effects, and subjects as a random effect, were performed separately for physically identical, same-name, and different responses. The critical $P \times P$ interaction for physically identical, same-name, and different pairs is depicted in Figure 4.

The only significant effect found for the physically identical pairs was the main effect of syllable type: VC syllables (307 msec) $\approx$ V syllables (312 msec) $<$ CV syllables (357 msec) $\approx$ CVC syllables (369 msec) [$F(3,135) = 6.71, p < .001$].

The results for same-category pairs were entirely different. Here, the $P \times P$ interaction attained significance [$F(1,135) = 5.16, p < .05$; Figure 4B]. Primed high-prototypical pairs were responded to faster (351 msec) than were unprimed pairs (418 msec). Computation of a planned t test confirmed that the difference was significant [$t(78) = 1.67, p < .05$]. Primed poor-example pairs were responded to slower than were unprimed pairs (493 msec vs. 444 msec), but the difference was not significant.

Figure 4. Mean response times in milliseconds for the priming $\times$ goodness level interaction effect for primed syllable matching under same-category instructions: (A) physically identical stimulus pairs, (B) same-category stimulus pairs, (C) different stimulus pairs. Full line = primed pairs; broken line = unprimed pairs.

Three other effects proved significant. First, good-example pairs were responded to faster than were poor-example pairs [379 msec vs. 462 msec; $F(1,135) = 10.82, p < .01$]. Second, the main effect of syllable type was significant. VC syllables were responded to faster than were CV, V, or CVC syllables [319 msec $<$ 454 msec $\approx$ 488 msec $\approx$ 444 msec; $F(3,135) = 8.27, p < .001$]. Finally, the priming $\times$ type interaction was

significant [$F(3,135) = 3.05$, $p < .05$]. A posteriori Tukey tests ($p = .01$) of the simple main effect revealed no significant differences within category types.

An ANOVA on pairs of different stimuli revealed two significant effects. First, the type $\times$ goodness level interaction was significant [$F(3,135) = 4.32$, $p < .01$]. A posteriori Tukey tests ($p < .01$) revealed no significant differences within category types. The priming $\times$ type interaction was also significant [$F(3,135) = 3.69$, $p < .05$]. A posteriori Tukey tests showed no differences within category types.

Discussion of Experiments 4A and 4B

Since a $P \times P$ interaction effect was obtained (in Experiment 4B), it can be inferred that the graded structure of syllable categories found in Experiment 2 (by asking the subjects to judge the goodness-of-example level of stimuli) was not an artifact due to either spurious prototypicality ratings or ratings determined by systematic covariates of the stimuli, but that it affected cognitive processing in a speech-relevant way. The goodness-of-example judgments of subjects for natural syllable categories correspond to processing differences between good and poor examples of a category.

However, contrary to the results of Experiments 3A and 3B with phoneme categories, the results of Experiments 4A and 4B revealed a $P \times P$ level interaction only for same-category pairs under same-category instruction. Physically identical pairs showed no priming $\times$ goodness level interaction, neither under physical identity instructions nor under same-category instructions. Only when the subjects were allowed to compare stimuli on the basis of more than purely perceptual information did priming facilitate response times, and even then only for same-category pairs.

A first conclusion is, of course, that syllable categories are different from phoneme categories. Since the representational content of syllable categories is not perceptually determined, as it is for phoneme categories, it has to be nonsensory. An obvious way to visualize such information is as an abstract rule based on phonological constraints dealing with the combination of phoneme categories (e.g., syllabification rules; Basbøll, 1988). However, this cannot explain why physically identical pairs under same-name instructions did not show a significant priming $\times$ goodness level interaction—physically identical pairs are also same-name pairs.

Moreover, the response times to physically identical pairs under same-name instructions (Figure 4A) were much faster than the response times to physically identical pairs under physical identity instructions (Figure 3A). Overall, the response time to physically identical pairs under same-name instructions was 336 msec, versus 504 msec for physically identical pairs under physical identity instructions. This seems to indicate that the representational format of syllables may be abstract in the sense specified above, but that it nevertheless contains a perceptual substratum. According to this view, the representational content of syllable categories could be the

specification of higher order transformations of specific perceptual information. An interesting picture emerges from this hypothesis. Under physical identity instructions, subjects have to compare the syllables in their detailed, low-level sensory realization, whereas under same-name instructions, the accessibility of the abstract, higher order perceptual information allows a very fast decision concerning identity—precisely because, for same-name matching, attention to concrete details of the syllable categories is not necessary here. The absence of a $P \times P$ interaction for these stimulus pairs is then explained by a floor effect based on the accessibility of information allowing subjects to decide early in the presentation of the second syllable of the stimulus pair that the two syllables belong to the same category. For physically different same-category pairs, the decision has to be postponed, allowing for the differential effect of priming on goodness level—as can be seen from Figures 4A and 4B, response times to primed good-example same-name pairs were comparable to response times to physically identical pairs. This hypothesis has implications both for phoneme and for syllable representations, as well as for further research. In the General Discussion section, these implications will be spelled out.

Some other effects also attained significance, of which only the type $\times$ goodness level interaction for physically identical pairs under physical identity instructions and the type $\times$ priming interaction for same-name pairs under same-name instructions need some comment, since these effects could in principle undermine the confidence in the critical results. However, since the three-way interactions were not significant, the $P \times P$ interaction can be interpreted in its own right (Keppel, 1982; Kirk, 1982; Winer, 1971).

In summary, the results of Experiments 4A and 4B show that the reliable prototypicality effects found in Experiment 2 refer to real speech-relevant differences between instances of syllable categories and that the representation of syllable categories has no simple perceptual content.

SUMMARY AND GENERAL DISCUSSION

In the series of experiments reported here, the primed matching task (Beller, 1971; Rosch, 1975a, 1975b) was used to investigate the speech unit hypothesis. The logic of the primed matching task allows one to differentiate directly the abstraction level of representations. If speech segments have graded structure, as can be expected from a prototype view of mental categories, priming should facilitate matching good-example pairs of a category, whereas it should have no, or possibly even an inhibitory, effect on matching poor-example pairs. Moreover, the results of Rosch (1975a, 1975b) showed that the primed matching task used with various matching instructions (physical identity or same name) could differentiate categories mentally represented at various levels of abstraction, ranging from entirely perceptual to entirely abstract. This is exactly what the speech unit hypothesis is

all about: Which speech segment has a more perceptually salient representational format—phoneme or syllable categories?

A direct prototypicality rating procedure was used in Experiments 1 and 2. Graded structure was shown to exist for both phoneme and syllable categories: Subjects could reliably assess a common graded structure for instances of phoneme and syllable categories. These ratings formed the basis of the four subsequent experiments, which investigated whether phoneme and syllable categories could be differentiated as far as abstraction level is concerned.

The results of Experiments 3A, 3B, 4A, and 4B revealed an interesting pattern of significant $P \times P$ interactions. Under both physical identity instructions and same-category instructions, a significant $P \times P$ interaction was obtained for physically identical pairs of phonemic categories: Response times to highly prototypical physically identical pairs were facilitated when primed. Poor-example physically identical pairs were unaffected by priming. Moreover, under same-category instructions, no $P \times P$ interaction was found for same-category pairs. On the other hand, for syllables under physical identity instructions, no $P \times P$ interaction was found. Even under same-category instructions, no $P \times P$ interaction was revealed for physically identical pairs. However, with same-category instructions, the interaction proved significant for same-category pairs.

In terms of the speech unit hypothesis, it can be concluded that phoneme categories are the basic units in speech perception. They are more perceptually based—hence, less abstract—than syllables. Moreover, because same-name pairs under same-name instructions showed no significant $P \times P$ interaction, phoneme representations seem to be entirely concrete. This in turn implies that phoneme categories form good patterns: The acoustic specification of a phonetic segment in the speech signal has to be such that a mental representation can be constructed from entirely perceptual characteristics. That is, the signal contains information allowing a stable perceptual description of phonetic segments by the perceptual system.

On the other hand, the results—a significant $P \times P$ interaction only for same-category pairs under same-category instructions—indicate that syllable representations are not in any direct way perceptually defined, but that they are characterized by more abstract information. However, the results for physically identical pairs under same-category instructions suggest that this abstract information still has a perceptual basis. The results for syllable categories show that a direct, low-level mapping from acoustic description to mental representation does not occur for these categories. Instead, they seem to be based on more abstract, higher order perceptual information that cannot be found directly in the acoustic signal, but which, for example, is defined over the low-level perceptual information specifying phoneme categories.

Although the experiments reported here did not aim at a detailed investigation of the representational content of

phoneme and syllable categories but only at a general distinction between a more concrete versus more abstract representational content, the specific pattern of results suggests an interesting view of the relation between phoneme and syllable representations. Given that phoneme categories are more basic than syllable categories, with syllable representations based on abstract but still (indirectly specified) perceptual information, syllable representations could possibly be understood as specifying the dynamics that arise from the realization of concatenated phonetic segments. Comparable suggestions have been made recently by Whalen (1989), describing syllables as imposing constraints on the shape of their constituent basic phonetic segments, and by Sussman (1984, 1986, 1989), whose neuronal model of speech sound development assumes that syllables are the basic organizational units for serial concatenation of sublexical elements (see also Browman & Goldstein, 1986, 1988a, 1988b, and Goldsmith, 1990, for similar models in a linguistic context). The information forming syllable representations is (1) more abstract than the information on which phoneme categories are based because it is structural, and (2) higher order perceptual because it specifies the dynamics of concatenating phonetic segments. This view has interesting implications for further research. In our laboratory, various methods are currently being explored to investigate the temporal relationship between phoneme and syllable perception in the light of this general view of phoneme and syllable representation.

Of course, further research with the primed matching task is needed to corroborate this interpretation. Since the research reported here is the first to my knowledge to use the primed matching task in studying the speech unit problem, it is important to assess the stability of the results that were found here. Some research possibilities can be pointed out. First of all, in Experiments 3A and 3B, vowels and consonants were treated as basically the same kind of category. While my results do not indicate any clear differences as far as the crucial priming $\times$ goodness level interaction is concerned, it has been known that differences do exist (see, e.g., Studdert-Kennedy, 1976, for a review). By using the primed matching task with vowels and consonants separately, this could be investigated more clearly. Second, prototypicality ratings were obtained by using only one speaker. However, it is possible that introducing talker variability would modify the results (e.g., Mullenix & Pisoni, 1990; Mullenix, Pisoni, & Martin, 1989). Finally, another interesting step will be the comparison of primed phoneme and syllable matching for categories that can be defined both as phoneme and as syllable categories (V syllables). By simply modifying the experimental instructions, it could be investigated if subjects respond differently depending on whether they are “forced” to treat categories as phonemic or syllabic.

Nevertheless, while other work with the primed matching task will have to be done before the results of the present research can take their full force, it does seem

clear that the primed matching task is indeed a potentially strong method for investigating the speech unit problem, and that it can offer a new perspective on this problem.

Finally, a metatheoretical implication of this research concerns the special-mechanisms debate in speech perception (Kuhl, 1987; Liberman & Mattingly, 1989). The theoretical and methodological underpinning of the primed matching task is a categorization perspective applicable to category representation in general. Graded structure is common to the representation of most categories (Medin & Barsalou, 1987). Finding that the general principles of this perspective can be used to study the speech unit problem bodes well for the integration of speech perception with the study of these general cognitive processes. The research reported here has shown that such an integration can elucidate some traditional problems in speech perception, such as the speech unit hypothesis.⁴

REFERENCES

- BASBØLL, H. (1988). Phonological theory. In F. J. Newmeyer (Ed.), *Linguistics: The Cambridge survey* (Vol. 1, pp. 192-215). Cambridge: Cambridge University Press.
- BELLER, H. K. (1971). Priming: Effects of advance information on matching. *Journal of Experimental Psychology*, **87**, 176-182.
- BOVENS, N., & BRYSAERT, M. (1990). IBM PC/XT/AT Turbo Pascal timing with extended resolution. *Behavior Research Methods, Instruments, & Computers*, **22**, 332-334.
- BROWMAN, C. P., & GOLDSTEIN, L. M. (1986). Towards an articulatory phonology. *Phonology Yearbook*, **3**, 219-252.
- BROWMAN, C. P., & GOLDSTEIN, L. M. (1988a). Articulatory gestures as phonological units. *Phonology Yearbook*, **5**, 201-251.
- BROWMAN, C. P., & GOLDSTEIN, L. M. (1988b). Some notes on syllable structure in articulatory phonology. *Phonetica*, **45**, 140-155.
- BURTON, M. W., BAUM, S. R., & BLUMSTEIN, S. E. (1989). Lexical effects on the phonetic categorisation of speech: The role of acoustic structure. *Journal of Experimental Psychology: Human Perception & Performance*, **15**, 567-575.
- CLEMENTS, G. N., & KEYSER, S. J. (1983). *CV Phonology: A generative theory of the syllable*. Cambridge, MA: MIT Press.
- COLLINS, B., & MEES, I. (1981). *The sounds of English and Dutch*. The Hague: Leiden University Press.
- CUTLER, A., MEHLER, J., NORRIS, D., & SEGUI, J. (1983). A language-specific comprehension strategy. *Nature*, **304**, 159-160.
- CUTLER, A., MEHLER, J., NORRIS, D., & SEGUI, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory & Language*, **25**, 385-400.
- CUTLER, A., MEHLER, J., NORRIS, D., & SEGUI, J. (1987). Phoneme identification and the lexicon. *Cognitive Psychology*, **19**, 141-177.
- CUTLER, A., MEHLER, J., NORRIS, D., & SEGUI, J. (1989). Limits on bilingualism. *Nature*, **340**, 229-230.
- CUTLER, A., & NORRIS, D. (1985). Juncture detection. *Linguistics*, **23**, 689-705.
- FOSS, D. J., & BLANK, M. A. (1980). Identifying the speech codes. *Cognitive Psychology*, **12**, 1-31.
- FOSS, D. J., & SWINNEY, D. A. (1973). On the psychological reality of the phoneme: Perception, identification, and consciousness. *Journal of Verbal Learning & Verbal Behavior*, **12**, 246-257.
- FRAUENFELDER, U. H., & SEGUI, J. (1989). Phoneme monitoring and lexical processing: Evidence for associative context effects. *Memory & Cognition*, **17**, 134-140.
- FRAUENFELDER, U. H., SEGUI, J., & DIJKSTRA, T. (1990). Lexical effects in phoneme processing: Facilitatory or inhibitory? *Journal of Experimental Psychology: Human Perception & Performance*, **16**, 77-91.
- GOLDSMITH, J. A. (1990). *Autosegmental and metrical phonology*. Oxford: Blackwell.
- GRIESER, D., & KUHL, P. (1989). Categorisation of speech by infants: Support for speech-sound prototypes. *Developmental Psychology*, **25**, 577-588.
- HEALY, A. F., & CUTTING, J. E. (1976). Units of speech perception: Phoneme and syllable. *Journal of Verbal Learning & Verbal Behavior*, **15**, 73-83.
- JAEGER, J. J. (1980). Testing the psychological reality of phonemes. *Language & Speech*, **23**, 233-253.
- JAEGER, J. J. (1986). Concept formation as a tool for linguistic research. In J. J. Ohala & J. J. Jaeger (Eds.), *Experimental phonology* (pp. 211-237). Orlando: Academic Press.
- JUSCZYK, P. (1986). Speech perception. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 2, pp. 27-1-27-57). New York: Wiley.
- KEPPEL, G. (1982). *Design and analysis*. Englewood Cliffs, NJ: Prentice-Hall.
- KIRK, R. E. (1982). *Experimental design: Procedures for the behavioral sciences*. Belmont: Brooks/Cole.
- KUHL, P. (1987). The special mechanisms debate in speech research: Categorisation tests on animals and infants. In S. Harnad (Ed.), *Categorical perception* (pp. 355-386). Cambridge: Cambridge University Press.
- KUHL, P. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, **50**, 93-107.
- KUHL, P., WILLIAMS, K., LACERDA, F., STEVENS, K., & LINDBLOM, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, **255**, 606-608.
- LAKOFF, G. (1986). *Women, fire, and dangerous things: What categories tell us about the nature of thought*. Chicago: University of Chicago Press.
- LIBERMAN, A., & MATTINGLY, I. (1989). A specialization for speech perception. *Science*, **243**, 489-494.
- MASSARO, D. (1972). Preperceptual images, processing time, and perceptual units in auditory perception. *Psychological Review*, **79**, 124-145.
- MASSARO, D. (1974). Perceptual units in speech recognition. *Journal of Experimental Psychology*, **102**, 199-208.
- MASSARO, D. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry*. Hillsdale, NJ: Erlbaum.
- MASSARO, D. (1989). Multiple book review of speech perception by ear and eye: A paradigm for psychological inquiry. *Behavioral & Brain Sciences*, **12**, 741-794.
- MCNEILL, D., & LINDIG, K. (1973). The perceptual reality of phonemes, syllables, words, and sentences. *Journal of Verbal Learning & Verbal Behavior*, **12**, 419-430.
- MEDIN, D., & BARSALOU, L. (1987). Categorisation processes and categorical perception. In S. Harnad (Ed.), *Categorical perception* (pp. 455-490). Cambridge: Cambridge University Press.
- MEHLER, J., DOMMERGUES, J. Y., FRAUENFELDER, U., & SEGUI, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning & Verbal Behavior*, **20**, 298-305.
- MEHLER, J., SEGUI, J., & FRAUENFELDER, U. (1981). The role of the syllable in language acquisition and perception. In T. Myers, J. Laver, & J. Anderson (Eds.), *The cognitive representation of speech* (pp. 295-305). Amsterdam: North-Holland.
- MELTZER, R. H., MARTIN, J. G., MILLS, C. B., IMHOFF, D. L., & ZOHAR, D. (1976). Reaction time to temporally-displaced phoneme targets in continuous speech. *Journal of Experimental Psychology: Human Perception & Performance*, **2**, 277-290.
- MILLER, J. L., CONNINE, C. M., SCHERMER, T. M., & KLUENDER, K. R. (1983). A possible auditory basis for internal structure of phonetic categories. *Journal of the Acoustical Society of America*, **73**, 2124-2133.
- MILLER, J. L., & VOLAITIS, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics*, **46**, 505-512.
- MILLS, C. B. (1980a). Effects of context on reaction time to phonemes. *Journal of Verbal Learning & Verbal Behavior*, **19**, 75-83.
- MILLS, C. B. (1980b). Effects of the match between listener expectations and coarticulatory cues on the perception of speech. *Journal of*

- Experimental Psychology: Human Perception & Performance*, 6, 528-535.
- MULLENIX, J., & PISONI, D. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics*, 47, 379-390.
- MULLENIX, J., PISONI, D., & MARTIN, C. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America*, 85, 365-378.
- NATHAN, G. (1986). Phonemes as mental categories. In *Proceedings of the 12th Annual Meeting of the Berkeley Linguistics Society* (pp. 212-223). Berkeley, CA.
- NORRIS, D., & CUTLER, A. (1988). The relative accessibility of phonemes and syllables. *Perception & Psychophysics*, 43, 541-550.
- NUNNALLY, J. C., JR. (1970). *Introduction to psychological measurement*. New York: McGraw-Hill.
- ODEN, G., & MASSARO, D. (1978). Integration of featural information in speech perception. *Psychological Review*, 85, 172-192.
- PISONI, D. B., & LUCE, P. A. (1987). Acoustic-phonetic representation in word recognition. In U. H. Frauenfelder & L. K. Tyler (Eds.), *Spoken word recognition* (pp. 21-52). Cambridge, MA: MIT Press.
- ROSCH, E. (1975a). Cognitive representation of semantic categories. *Journal of Experimental Psychology: General*, 104, 192-233.
- ROSCH, E. (1975b). The nature of mental codes for color categories. *Journal of Experimental Psychology: Human Perception & Performance*, 1, 303-322.
- ROSCH, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27-48). Hillsdale, NJ: Erlbaum.
- RUBIN, P., TURVEY, M. T., & VAN GELDER, P. (1976). Initial phonemes are detected faster in spoken words than in spoken nonwords. *Perception & Psychophysics*, 19, 394-398.
- SAMUEL, A. (1982). Phonetic prototypes. *Perception & Psychophysics*, 31, 307-314.
- SAMUEL, A. (1989). Insights from a failure of selective adaptation: Syllable-initial and syllable-final consonants are different. *Perception & Psychophysics*, 45, 485-493.
- SAVIN, H. B., & BEVER, T. G. (1970). The nonperceptual reality of the phoneme. *Journal of Verbal Learning & Verbal Behavior*, 9, 295-302.
- SCHARF, B., & BUUS, S. (1986). Audition: I. Stimulus, physiology, thresholds. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance: I. Sensory processes and perception* (pp. 14-1-14-71). New York: Wiley.
- SCHROUT, P. E., & FLEISS, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86, 420-428.
- SEGUI, J. (1983). The syllable: A basic perceptual unit in speech processing? In H. Bouma & D. Bouwhuis (Eds.), *Attention and performance X. Control of language processes* (pp. 165-181). Hillsdale, NJ: Erlbaum.
- SHAND, M. A. (1976). Syllabic vs. segmental perception: On the inability to ignore "irrelevant" stimulus parameters. *Perception & Psychophysics*, 20, 430-432.
- STINE, W. W. (1989). Interobserver relational agreement. *Psychological Bulletin*, 106, 341-347.
- STUDDERT-KENNEDY, M. (1976). Speech perception. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics* (pp. 243-293). Springfield, IL: Thomas.
- SUSSMAN, H. (1984). A neuronal model for syllable representation. *Brain & Language*, 22, 167-177.
- SUSSMAN, H. (1986). A neuronal model of vowel normalisation and representation. *Brain & Language*, 28, 12-23.
- SUSSMAN, H. (1989). Neural coding of relational invariance in speech: Human language analogs to the barn owl. *Psychological Review*, 96, 631-642.
- SWINNEY, D. A., & PRATHER, P. (1980). Phonemic identification in a phoneme monitoring experiment: The variable role of uncertainty about vowel contexts. *Perception & Psychophysics*, 27, 104-110.
- UIT DEN BOOGAERT, P. (1975). *Woordfrequenties in geschreven en gesproken Nederlands*. Utrecht: Scheltema & Hoekema.
- VAN DER HULST, H. (1984). *Syllable structure and stress in Dutch*. Dordrecht: Foris.
- WHALEN, D. (1989). Vowel and consonant judgments are not independent when cued by the same information. *Perception & Psychophysics*, 46, 284-292.
- WINER, B. (1971). *Statistical principles in experimental design*. New York: McGraw-Hill.
- WOOD, C. C., & DAY, R. S. (1975). Failure of selective attention to phonetic segments in consonant-vowel syllables. *Perception & Psychophysics*, 17, 346-350.
- ZWITSERLOOD, C. M. (1989, May 19). *Over de rol van de syllabe bij het waarnemen van gesproken woorden*. Lecture held at the Institute for Perception Research, Eindhoven, The Netherlands.

NOTES

1. According to Cutler et al. (1986), languages with variable syllable structure and unclear syllable boundaries should show no evidence of listeners' using the syllable as a segmentation unit. Recently, however, results for Dutch were reported by Zwitserlood (1989). Dutch has both clear and unclear boundaries. In each case, Zwitserlood obtained clear syllable effects. Therefore, the interpretation of language-specific syllable effects in the monitoring task remains unclear.

2. Computing whether or not the difference between these means is significant is not necessary. The interaction effect between priming and goodness level was predicted a priori. This means that if the interaction is significant and if the interaction conforms to the prediction, nothing more has to be computed (Keppel, 1982; Kirk, 1982; Winer, 1971). Planned *t* tests have been computed for illustration only.

3. A purely qualitative corroborating argument—without statistical foundation—comes from splitting the three-way interaction according to stimulus type. Under physical identity instructions, primed and unprimed good-example consonant pairs were matched equally fast (453 msec vs. 445 msec), and primed poor-example consonant pairs were matched slower than were unprimed pairs (528 msec vs. 380 msec), primed good-example vowel pairs were matched faster than were unprimed pairs (288 msec vs. 412 msec), and primed poor-example vowel pairs were matched somewhat faster than were unprimed pairs (334 msec vs. 374 msec). Under same-name instructions, primed good-example consonant pairs were matched faster than were unprimed pairs (367 msec vs. 416 msec), and primed poor-example pairs were matched slower than were unprimed pairs (451 msec vs. 384 msec); primed good-example vowel pairs were matched faster than were unprimed pairs (328 msec vs. 431 msec), and primed poor-example vowel pairs were matched about as fast as were unprimed pairs (378 msec vs. 409 msec). These patterns by stimulus type conform to the overall priming $\times$ goodness level pattern, with graded structure of consonant categories showing more clearly in the poor-example region, and graded structure of vowel pairs showing more clearly in the good-example region. So, any differences between consonant and vowel categories could be related to, for example, the specific location of prototypicality effects in a vowel and consonant "space." This is, however, speculation. It is clear, though, that the crucial priming $\times$ goodness level interaction can be interpreted with confidence.

4. This conclusion agrees well with the research of, for example, Jaeger (1980, 1986) and Nathan (1986), which has stressed the usefulness of concept formation and categorization methods within the domain of phonetics and phonology.

APPENDIX A

Stimulus Words Used in Experiments 3A and 3B

The stimulus pairs for each category are presented in the following order: (1) good-example physically identical pair, (2) poor-example physically identical pair, (3) good-example different pair, (4) poor-example different pair, (5) good-example same-name pair, and (6) poor-example same-name pair. Stress is indicated by an apostrophe after the stressed syllable.

/s/ (1) siroop'-siroop' (2) schijn'-schijn' (3) siroop'-kaars' (4) schijn'-exa'men (5) siroop'-signaal' (6) schijn'-scep'per
/r/ (1) reak'tie-reak'tie (2) ramp'-ramp' (3) reak'tie-arrest'

- (4) ramp'-aal'fuik (5) reak'tie-rij'den (6) ramp'-rap'
 /k/ (1) kaars'-kaars' (2) kwe'zel-kwe'zel (3) kaars'-siroop'
 (4) kwe'zel-elatiet' (5) kaars'-ka'mer (6) kwe'zel-koe'gang
 /e:/ (1) eer'zucht-eer'zucht (2) elatiet'-elatiet' (3) eer'zucht-aard'bei (4) elatiet'-schijn' (5) eer'zucht-e'ven (6) elatiet'-ejek'tor
 /ɛ:/ (1) enzy'me-enzy'me (2) exa'men-exa'men (3) enzy'me-reak'tie (4) exa'men-arbitrair' (5) enzy'me-eldora'do (6) exa'men-enkadre'ren
 /a/ (1) aard'bei-aard'bei (2) aal'fuik-aal'fuik (3) aard'bei-enzy'me (4) aal'fuik-kwe'zel (5) aard'bei-aar'zeling (6) aal'fuik-aan'baksel
 /a/ (1) arrest'-arrest' (2) arbitrair'-arbitrair' (3) arrest'-eer'zucht (4) arbitrair'-ramp' (5) arrest'-arduin' (6) arbitrair'-annule'ren

APPENDIX B

Stimulus Words Used in Experiments 4A and 4B

The stimulus pairs for each category are presented in the following order: (1) good-example physically identical pair, (2) poor-example physically identical pair, (3) good-example different pair, (4) poor-example different pair, (5) good-example same-name pair, and (6) poor-example same-name pair. Stress is indicated by an apostrophe after the stressed syllable.

- /a/ (1) ateis'me-ateis'me (2) are'ool-are'ool (3) ateis'me-o'lie (4) are'ool-op'bellen (5) ateis'me-atopie' (6) are'ool-achil'les
 /o:/ (1) o'lie-o'lie (2) o'ryx-o'ryx (3) o'lie-op'treden (4) o'ryx-as'zeef (5) o'lie-o'gen (6) o'ryx-oriënt'
 /na:/ (1) na'komer-na'komer (2) na'loot-na'loot (3) na'komer-

ateis'me (4) na'loot-kop'blok (5) na'komer-na'galm (6) na'loot-na'lente

/be/ (1) bekij'ken-bekij'ken (2) beroep'-beroep' (3) bekij'ken-parabool' (4) beroep'-vak'kast (5) bekij'ken-bete'kenen (6) beroep'-beroemd'

/pa/ (1) parabool'-parabool' (2) paret'-paret' (3) parabool'-bekij'ken (4) paret'-beroep' (5) parabool'-para'bel (6) paret'-palindroom'

/as/ (1) as'kleur-as'kleur (2) as'zeef-as'zeef (3) as'kleur-uit'gever (4) as'zeef-are'ool (5) as'kleur-as'kern (6) as'zeef-as'gat

/uit/ (1) uit'gever-uit'gever (2) uit'deling-uit'deling (3) uit'gever-as'kleur (4) uit'deling-na'loot (5) uit'gever-uit'roeien (6) uit'deling-uit'braak

/op/ (1) op'treden-op'treden (2) op'bellen-op'bellen (3) op'treden-na'komer (4) op'bellen-o'ryx (5) op'treden-optimist' (6) op'bellen-op'breken

/kop/ (1) kop'stuk-kop'stuk (2) kop'blok-kop'blok (3) kop'stuk-vet'stof (4) kop'blok-paret' (5) kop'stuk-kop'zorg (6) kop'blok-kop'bout

/vɛt/ (1) vet'stof-vet'stof (2) vet'je-vet'je (3) vet'stof-vak'verbond (4) vet'je-uit'deling (5) vet'stof-vet'laag (6) vet'je-vet'mot

/vak/ (1) vak'verbond-vak'verbond (2) vak'kast-vak'kast (3) vak'verbond-kop'stuk (4) vak'kast-vet'je (5) vak'verbond-vak'opleiding (6) vak'kast-vak'kring

(Manuscript received September 23, 1991;
 revision accepted for publication November 17, 1992.)