

Cross-series adaptation using song and string

ROBERT E. REMEZ

Indiana University, Bloomington, Indiana 47405

JAMES E. CUTTING

Cornell University, Ithaca, New York 14853

and

MICHAEL STUDDERT-KENNEDY

*Department of Communication Arts and Sciences, Queens College of CUNY, Flushing, New York 11367
and Haskins Laboratories, New Haven, Connecticut 06510*

Cross-series adaptation using song and string

ROBERT E. REMEZ

Indiana University, Bloomington, Indiana 47405

JAMES E. CUTTING

Cornell University, Ithaca, New York 14853

and

MICHAEL STUDDERT-KENNEDY

Department of Communication Arts and Sciences, Queens College of CUNY, Flushing, New York 11367
and Haskins Laboratories, New Haven, Connecticut 06510

The acoustic-auditory feature "risetime" has been claimed to underlie both the phonetic *affricate-fricative* distinction and the nonphonetic *plucked-string/bowed-string* distinction. We used the perceptual adaptation technique to determine whether the risetime differences of the [dʒa]-[ʒa] distinction would therefore be registered by the same mechanism that mediates risetime differences for the plucked-bowed distinction. Two continua were used, one of digitally modified natural speech and one of synthetic violin sounds, in which the risetime was varied across each set of tokens from 0 msec to 80 msec in steps of 10 msec. The speech was sung and the violin notes were synthesized with the same fundamental frequency, 294 Hz. Adaptation of the category boundaries was observed only when speech adaptors were tested with the speech continuum and when violin adaptors were tested with the violin continuum. When cross-series tests were performed (violin adaptors tested with the speech series, and speech adaptors tested with the violin series), no effect of adaptation was observed. This finding indicates that these speech and violin sounds, despite obvious acoustic similarities, do not share the same feature detectors.

Is speech perception merely an auditory process? The discovery that different kinds of synthetic violin sounds are perceived categorically (Cutting & Rosner, 1974) seemed to suggest that it might be. Categorical perception had previously been thought to result from a decoding process of the speech perception system (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). Phonetic decoding of acoustic signals was believed to be abstractly keyed to articulatory patterns of phonetic production (Cooper, Liberman, Harris, & Grubb, 1958). In the earliest form of the hypothesis, speech segments produced by graded motor signals (e.g., vowels) were said to be perceived in a continuous fashion, in contrast with those produced by discrete motor commands (e.g., consonants), which were said to be perceived cate-

gorically. Though subsequent study of articulator coordination and motor control ruled out literal parity of production and perception (MacNeilage, 1970), categorical perception continued to be an important criterion for separating speech perception from general auditory perception (Liberman, 1970). However, when Cutting and Rosner (1974) reported that a synthetic stimulus continuum of violin sounds ranging from *plucked string* to *bowed string* was categorically perceived, they clearly eliminated categorical perception as one of the definitive qualities of perception in the speech mode. This result, in combination with other similar findings (e.g., Locke & Kellar, 1973; Miller, Weir, Pastore, Kelly, & Dooling, 1976; Pisoni, 1977), intensified the appeal of explaining speech perception in terms of general auditory (lower level) mechanisms.

One potential account of the categorical perception of violin "articulation" was that the distinction between plucked string and bowed string was actually mediated by mechanisms intrinsic to the speech system. Cutting and Rosner (1974) considered this possibility because they saw that the acoustic basis for perceiving these particular violin sounds is very nearly identical to that of the phonetic *affricate-fricative* distinction (Gerstman, 1957). Both plucked-bowed and *affricate-*

This research was supported by a grant from NICHD (HDO 1994-14) to Haskins Laboratories and a grant from NIMH (MH 32848-01) to the first author. We wish to thank David B. Pisoni for extending to us the use of the computer facilities at the Speech Perception Laboratory, Indiana University (supported by NIMH Grant MH 24027-05 and by NIH Grant NS 12174-03). We are also grateful to Linda B. Smith for her advice and comments on our manuscript. Address correspondence to Robert E. Remez, Department of Psychology, Barnard College, New York, New York 10027.

fricative may be distinguished minimally by the rate of onset of acoustic energy, the risetime of the amplitude envelope. When risetime is relatively brief, shorter than 40 msec, listeners hear a plucked string, or an affricate consonant; when risetime is great, longer than 40 msec, listeners hear a bowed string, or a fricative (cf. Cutting & Rosner, 1976). However, Cutting and Rosner did not defend this proposal. Instead, they favored a general auditory mechanism that might serve both phonetic *and* violin distinctions by tracking amplitude changes in the acoustic signal. This auditory explanation seemed more plausible than the proposal that the speech system had mistakenly processed the synthetic violin sounds, primarily because it was not obvious why the speech decoder would be misapplied to these particular nonphonetic patterns. In addition, such an explanation in the case of violins would be inconsistent with supposed biological advantages of categorical perception. If the biologically determined speech code is unique among the acoustic patterns which engage the auditory system, then the speech processor has a very easy job of telling speech sounds from nonspeech (Liberman, Mattingly, & Turvey, 1972); it must merely listen only for stimuli of the appropriate structure. From this standpoint, it would make little sense to argue that the stable, efficient speech system would be easily fooled into "phonetically" processing the unequivocally nonspeech violin sounds. In short, Cutting and Rosner concluded that the perception of violin sounds is probably not accomplished by the speech perception mechanism.

The goal of Cutting, Rosner, and Foard (1976), then, was to determine a possible sensory basis of categorical perception (as opposed to the motoric rationale offered by Liberman et al., 1967), and they took the pattern recognition scheme of opponent-related feature detectors introduced by Eimas and Corbit (1973) as their model. But, while Eimas had originally conceived the detectors to be phonetic, Cutting et al. saw that *auditory* detectors would be required to serve parsimoniously for both speech *and* violin sounds. Had the violin sounds not been categorically perceived, the phonetic detector model could have trivially accommodated nonspeech phenomena at a lower, auditory level. However, since the plucked-bowed distinction was categorical, it appeared that the phenomenon of categorical perception should be recast nonphonetically, in more general auditory terms. To the degree that adaptation provides a means of establishing detector sensitivities (cf. Remez, 1979), the demonstration by Cutting et al. of adaptation effects on identification functions for synthetic violin sounds argued that categorical perceptions of nonspeech sounds, no less than speech sounds, might be mediated by the same set of auditory detectors.

Cutting et al. used their violin test series in two adaptation conditions; both conditions revealed significant adaptation effects. In the *within-series* condi-

tion, the adaptors were drawn from the test continuum; the adaptor was either the plucked endpoint or the bowed endpoint. In *selective* adaptation tests, adaptors were not drawn from the continuum but were fashioned to fatigue the specific risetime detectors thought to mediate perception of the test stimuli. Listeners were found in these cases to adapt to risetime selectively when adaptor and test series differed in waveform (sinusoid vs. sawtooth) or in fundamental frequency, although boundary shifts in those cases were diminished relative to within-series shifts. While the experiment did not directly test the hypothesis that a single auditory mechanism sensitive to risetime was responsible for the analysis of plucked-bowed strings and affricate-fricative consonants, it did reveal that a style of perceptual analysis similar to that proposed for speech appeared to operate at an auditory level for nonspeech sounds. The implication to be drawn from the adaptation results of Cutting et al. is clear: Auditory analyzers alone might suffice for the categorical perception of both speech and violin sounds alike.

The present experiment uses the adaptation paradigm in a direct test of the hypothesis that a single set of auditory analyzers tuned to amplitude risetime serves both speech and nonspeech perception. To perform this test, two stimulus series were fashioned, one of synthetic plucked-to-bowed violin sounds, the other of computer-modified natural affricate-to-fricative consonant-vowel syllables. These series were as similar as possible in the onset portions critical to the respective perceptual distinctions. In the speech case, the continuum ranged from [dʒa] (affricate; as in "judge") to [ʒa] (fricative; as in "vision") (after Gerstman, 1957); in the violin series, the synthetic continuum ranged from the sound of a plucked string to that of a bowed string (after Cutting & Rosner, 1974). Our test consisted of two parts: In the first, adaptation caused by consonant endpoint or violin endpoint adaptors was measured by using the stimulus test series of the same type as the adaptor. This is the *within-series* test. In the second, adaptation caused by speech adaptors was measured with the violin test series, that caused by violin adaptors measured with the speech series. This is the *cross-series* test. Our purpose was to establish whether an amplitude risetime detector fatigued by a nonspeech stimulus would effect a perceptual change in a speech test series, and vice versa. The demonstration of cross-series adaptation in this circumstance would be strong evidence for a common analytic mechanism underlying the categorical perception of speech and nonspeech sounds.

METHOD

Subjects

Eight volunteers served as listeners. All were undergraduate students enrolled at the University of Connecticut. They were right-handed native speakers of English; none had a history of speech

disorder or hearing impairment, nor did any confess more than a casual familiarity with musical instruments of the violin family. They were paid for their participation.

Stimuli

Two stimulus continua were employed, one of synthetic violin sounds and the other of computer-modified natural speech. The violin series was one of the nine-item sets of Cutting and Rosner (1974), in which the amplitude risetime of a 294-Hz sawtooth wave was varied from 0 msec to 80 msec in steps of 10 msec. Overall duration varied from 1,020 to 1,100 msec. The matching speech series was generated from a CV syllable, [ʒa], sung by one of the authors (J.E.C.) in falsetto register at 294 Hz. The syllable was recorded on audiotape, low-pass filtered at 5 kHz, sampled at 10 kHz, and digitized on the Haskins Laboratories pulse-code modulation system (Cooper & Mattingly, 1969). It was then modified by multiplying the digital record under computer control to produce a series of sounds that ranged from [dʒa] to [ʒa] by iterative adjustments in the duration of linear amplitude risetime.¹ The nine-item continuum varied from 0 to 80 msec of risetime; the duration of the voiced frication in the initial portion of each token covaried with risetime from 70 to 150 msec (after Gerstman, 1957). Waveforms of representative tokens are presented in Figure 1. Test sequences were recorded on audiotape by digital-to-analog conversion and were presented binaurally to listeners via Crown 820-144 playback through TDH-39 earphones at a comfortable level of about 72 dB SPL.

Procedure

Eight tests were presented on separate days over the course of 2 weeks; four were of within-series adaptation, four of cross-series. In the within-series conditions, the test continuum and adaptor were of the same kind, both either speech or violin. In cross-series tests, the adaptor was of one kind and the test series was of the

other. Subjects were tested in two groups of four. One group was tested on within-series first, cross-series second; the other group was tested in the reverse order.

A brief practice sequence of four presentations of each day's test series endpoints began the testing session. Following this was a baseline test of unadapted identification of the relevant stimulus series, a randomized 90-item sequence of the nine tokens repeated 10 times each. Three seconds of silence separated trials. After a short intermission, adapted identification was measured, following the procedure of Cutting et al. (1976). This consisted of 12 blocks of trials of alternating adaptor and identification portions. In the first block there were 100 adaptor repetitions followed by seven identification trials drawn from the nine-item continuum. The remaining 11 blocks were similar to the first, but with only 50 adaptor repetitions. There were 600 msec of silence between successive adaptors, 2 sec between adaptor block and identification block, 2 sec between successive identification trials, and 5 sec between the end of the identification block and the following adaptor block. Six observations per listener were obtained for items with risetimes of 0, 10, 70, and 80 msec, and 12 observations for those with risetimes of 20, 30, 40, 50, and 60 msec. Responses were scored on a prepared answer sheet as "P" (plucked), "B" (bowed), "J" ([dʒa]), or "Z" ([ʒa]).

RESULTS AND DISCUSSION

For each within- and cross-series adaptation test, we calculated two scores per subject: one for baseline identification and one for adapted identification. In each of the eight conditions, the proportion of "long rise" responses ([ʒa] or bowed) was compared in baseline and adapted tests by a *t* test for paired observa-

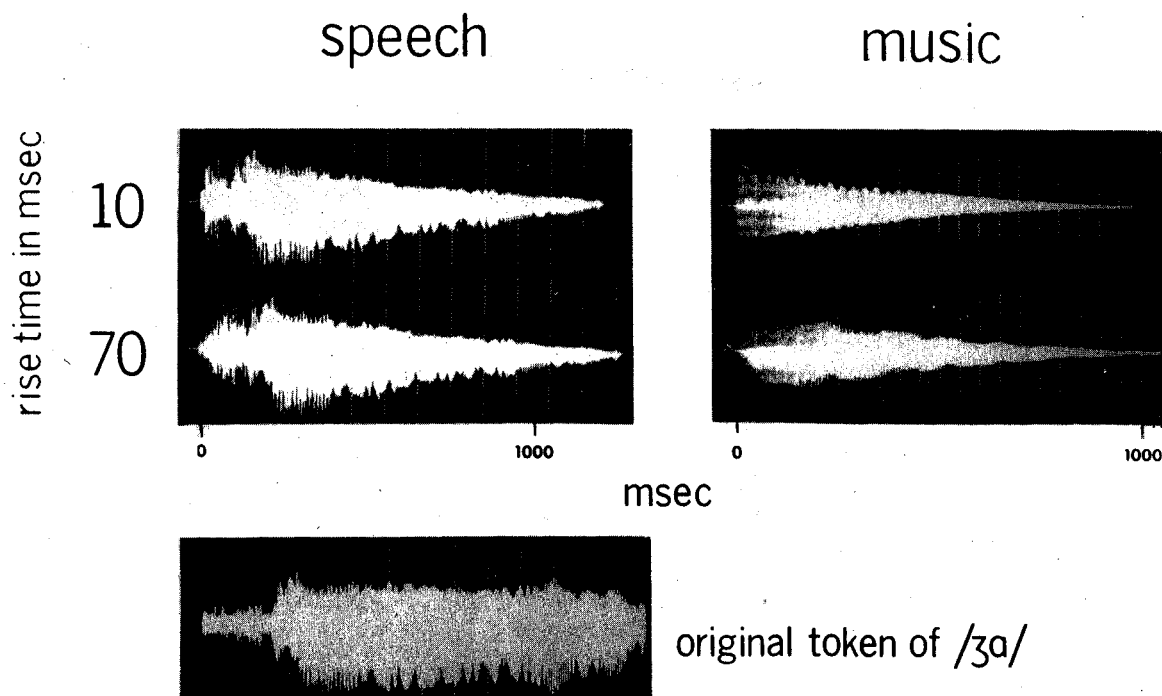


Figure 1. Top: Waveforms of representative tokens. Left: Speech, [dʒa] (above) and [ʒa] (below). Right: violin, plucked string (above) and bowed string (below). Bottom: Original [ʒa] utterance.

Table 1
Summary of Within-Series and Cross-Series Adaptation

Condition	Baseline – Adapted ID	t
[dʒa] on speech	.122	4.798**
[ʒa] on speech	-.089	3.927**
plucked on violin	.052	3.260*
bowed on violin	-.097	4.140**
[dʒa] on violin	.002	.072†
[ʒa] on violin	-.016	.588†
plucked on speech	.002	.154†
bowed on speech	.007	.364†

Note—Identification score (ID) is the proportion of long-rise responses (see text). ** $p < .005$. * $p < .01$. † $p > .1$.

tions. Within-series differences between baseline and adapted identification were statistically significant, but there were no significant differences between baseline and adapted identification in cross-series tests (violin adaptors tested with the speech series, and speech adaptors tested with the violin series). Table 1 summarizes the results.

Our experiment reveals that test continuum and adaptor must share more than merely the acoustic attributes of amplitude envelope and fundamental frequency to show cross-series adaptation effects. This finding is consistent in principle with the prior one of Cutting et al. (1976), that in cross-series adaptation the magnitude of the effect was a function of the auditory attributes that the test series shared with the adaptor. However, since the attribute of risetime was always common to adaptor and test series in their study, we had supposed that sounds sharing amplitude risetime might be mutually effective adaptors, regardless of the other auditory attributes distinguishing them. Risetime, as a perceptual commodity, might then be parsimoniously assigned to a single pair of detectors.

The outcome of the present experiment refutes this hypothesis. In our test, cross-series adaptors and test continua differed neither in amplitude risetime nor in fundamental frequency, but did differ in spectral composition, and this spectral difference (or, in other words, the apparent difference in sound source) was sufficient to prevent cross-adaptation of the two series. Evidently, speech and violin sounds do not share a single 0-80 msec pair of risetime detectors. By extrapolation, we may conclude that the categorical perception of the affricate-fricative and plucked-bowed distinctions does not rely on the detection of risetime by a single auditory pattern analyzer.

In this light, an earlier report by Diehl (1976) is puzzling. Diehl was testing the generality of the *stop* vs. *continuant* opposition using nonspeech sounds as adaptors. He found that a plucked adaptor was effective on a [ba]-[wa] continuum, as if it shared the property "stop" with [ba], although a similar test using a bowed adaptor showed no adaptation effect. What is even more curious is that the [ba] and the

plucked sound that had similar effects as adaptors were not similar in amplitude envelope, and that the [ba]-[wa] distinction was not cued by risetime. Finally, the frequency of the violin adaptors was 440 Hz, while the speech sounds presumably had a fundamental at least two octaves lower. In short, the cross-series adaptation reported by Diehl for violin and speech sounds, not sharing risetime, fundamental frequency, or spectral composition, must surely be the most abstract instance of adaptation ever observed. Whatever the critical attribute may have been, it was plainly not the abstract *stop* vs. *continuant* opposition, as Diehl claimed, since our [dʒa], which did not cross-adapt, is a proper stop and [ʒa] is a proper continuant.

A related set of findings was reported recently by Samuel and Newport (1979), who replicated Diehl's earlier puzzling result. They used two test series, [ba] to [wa] and [tʃa] to [ʃa], and four nonspeech adaptors, two periodic and two aperiodic, of short and long risetime each. The periodic nonspeech adaptor pair were plucked and bowed violin sounds which differed in fundamental period from the speech series, again, presumably, by two octaves. The aperiodic adaptor pair were broad-band noise patterns with digitally shaped envelopes; they did not sound like whispered speech (Samuel, Note 1). Each of the four adaptors was tested on both speech continua. The surprising findings in this study were that the periodic short risetime adaptor acted, again, as if it shared a property with [ba] but not with [tʃa], and the long risetime aperiodic adaptor acted as if it shared a property with [ʃa] but not with [wa]. Confronted by this set of asymmetrical risetime adaptation effects, the authors posited a variation of the original detector model of Eimas; they suggested that periodic waveforms were mediated by a single detector sensitive only to sharp envelope discontinuities, while aperiodic waveforms were mediated by a single detector sensitive only to gradual energy onset. However, one test which Samuel and Newport did not perform is crucial to supporting this conclusion. In order to verify the characterizations of the two auditory detectors, speech adaptors must also be effective when the test series is nonspeech. In other words, the proposal of auditory, periodicity-labeled, asymmetrically tuned risetime detectors is warranted only if the effects of speech adaptors on nonspeech test continua are shown to correspond to the effects of nonspeech adaptors on speech test continua.

In addition, our present test failed to confirm this hypothesis. By the criteria developed by Samuel and Newport, we would predict that only the plucked string and [dʒa] adaptors are effective, although each should produce adaptation on both test continua. That is because both the speech and the violin continua were periodic (at the same fundamental frequency, 294 Hz). The prediction was incorrect. More-

over, their conclusions are open to criticism independent of our disconfirming cross-series data. The detectors they have proposed account only for the success of plucked adaptors concomitant with the failure of bowed adaptors. Although this outcome occurred in their test, their conclusion is overly general; a periodic risetime detector that asymmetrically prefers sharp amplitude discontinuities would not explain the original symmetrical within-series and selective risetime adaptation effects of Cutting et al. (1976) or the symmetrical within-series effects noted here. Our data may therefore be taken to question the usefulness of the modification of the detector model endorsed by Samuel and Newport.²

GENERAL DISCUSSION

Our finding that adaptation effects are segregated by apparent sound source has a precedent. Faced with an analogous result, of fundamental-frequency-contingent phonetic adaptation, Ades (1977) had suggested that multiple sets of auditory detectors might exist, each assignable to an individual talker. In our case, we might extend this source-assignment notion to *sound* sources more generally, whether sources of phonetic segments or not. For instance, risetime detectors might be basic ingredients in auditory detector ensembles; speech and violin notes, produced by different sound sources, would then be mediated perceptually by separate but equal ensembles, one assigned to each sound source. This vague notion of reduplicated auditory detectors will handle our result, but it is both ad hoc and inelegant, particularly in view of the current controversy over the status of feature detectors in speech perception.

The short history and the interpretive difficulties of adaptation studies using phonetic materials have been reviewed by Cooper (1975) and by Eimas and Miller (1978). To the difficulties exposed by these authors, we may add the following: (1) Detectors conceived phonetically fail to predict the occasions on which adaptation does and *does not* occur; (2) detectors conceived auditorily have so proliferated as to eliminate the appealing simplicity of the original model—to derive perceptual categories in the auditory version of the model requires cumbersome and implausible detector interactions; and (3) the adaptation test itself may not be appropriate if the aim is to catalogue fixed preferences and sensitivities of analytic channels (see Diehl, Elman, & McCusker, 1978; Remez, 1979; Simon & Studdert-Kennedy, 1978). Whether the theory is largely false (as claimed by Diehl et al., 1978; Remez, 1980; Simon & Studdert-Kennedy, 1978) or is minimally erroneous (as claimed by Cooper, 1979; Eimas & Miller, 1978; Samuel & Newport, 1979) is a matter that has received much attention lately. However, it is obvious that with no explicit use for distinctive feature theory possible,

and with no manageable set of auditory templates in hand, the adaptation studies of phonetic stimuli can have little to say about the classic problems in speech perception (i.e., recognition, segmentation, normalization, etc.). We therefore question the worth of explaining our current finding within the feature detector model.

Furthermore, we find an independent dismissal of feature detector accounts in a recent study quite analogous to our own (Pisoni, in press). Pisoni observed no cross-series adaptation effects for speech and nonspeech stimulus sets that were equivalent on an acoustic dimension sometimes taken to be perceptually critical in speech (cf. Lisker & Abramson, 1964, 1971). The speech set was one of *voicing*, and the nonspeech set one of two-component tones which onset at differing degrees of asynchrony. As in the case of Cutting et al. (1976) and the present experiment, the acoustic criteria for the speech and nonspeech distinctions were apparently the same. The acoustic parameter which served to distinguish *coincident* from *delayed* voicing onset also served to distinguish the "categories" of the nonspeech set; this was the temporal relation of the onsets of the individual spectral components. In adaptation tests, the two-tone continuum exhibited within-series adaptation, as did voicing, which is well known to undergo adaptation (e.g., Eimas & Corbit, 1973), but neither series exhibited cross-series effects. Thus, identical temporal structure did not guarantee cross-adaptation.

Rather than invent a new pair of detectors to handle the finding, and rather than take the tack of Ades (1977) by inventing an entirely new *set* of detectors assigned ad hoc to nonspeech tones, Pisoni preferred to argue that temporal order judgments of any class might be limited by a general constraint of the auditory system unspecific to particular detectors. The amplitude risetime parameter that we have studied for violin and speech sounds may reflect a related general auditory limitation, one not specific to any pair of phonetic or auditory detectors. If the auditory system has a characteristic rate of temporal integration throughout its frequency range, for instance, then the transduction of every sound impinging on the system would reflect that limitation in resolution. In some circumstances, we may expect a perceptual distinction to hinge on differential effects of this auditory limitation. But whether or not the product of the general auditory constraint is perceptually distinctive in one or another instance, we should not assume that the mechanism responsible is a single pair of detectors or even a pair of ensembles of detectors. Were that the case, our closely matched set of speech and violin sounds would have adapted each other, for they were as similar as possible within the requirement that they differ in their apparent sources, one a talker, the other a violin.

It may also be valuable, finally, to consider the origins of the violin perceptual categories which appear coincidentally to be similar in acoustic time course to sounds of speech. Recently, Remez (1978) has argued that violins are capable of categorical and continuous modes of sound *production*, although this distinction is a mechanical one unlike the neurophysiological distinction once claimed in the case of speech (Cooper et al., 1958; Liberman et al., 1967; cf. Stevens, 1972). For example, the productive distinction *sul ponticello-sul tasto* is acoustically correlated with differences in spectral envelope; *sul ponticello* (played with the bow near the bridge) has a shallower rolloff, that is, more energy in the higher harmonics, than does *sul tasto* (played with the bow near the fingerboard), all other things being equal (Schelleng, 1973). Given constant bow force, then, this dimension of production is continuously variable in infinitely small steps as the point of contact of bow and string is moved from the bridge toward the fingerboard, or vice versa. Compare this with the *pizzicato-arco* (plucked-bowed) distinction. These terms name two quantal alternatives of sound production. Bowing involves what Schelleng (1973) has called the "stick-slip" interaction of string and bow, in which force is applied to the string in a relatively sustained manner. Plucking, on the other hand, involves a single loading of the string when it is retracted by a finger or other grasping agent (such as a plectrum in a harpsichord). Once loaded, the string is released. The differences in production are correlated acoustically with the risetime differences we have been discussing. Unlike the gradual change from *sul ponticello* to *sul tasto*, the distinction between *pizzicato* and *arco* is categorical mechanically; there is no mechanical gradient between plucking a string and bowing it. Listeners who perceive the mechanical events of instrumental sound production should, by this line of reasoning, perceive the plucked-bowed distinction categorically. Perhaps the finding of violin categoricity that has inspired the studies by Cutting and his colleagues is as much a function of the perception of *intrinsically categorical* mechanical events as it is the function of the auditory resolution of risetime.

To summarize, our experiment used a cross-series adaptation test to determine whether risetime of amplitude envelope is mediated by a single auditory mechanism in both phonetic and nonphonetic circumstances. Although we found within-series effects for speech and violin sounds, cross-series tests did not reveal adaptation. Thus, stimuli matched for fundamental frequency and amplitude risetime apparently do not share feature detectors—if we judge by the conventional criteria for establishing shared detectors. This conclusion in no way implies that detectors of any kind exist. We further conclude that the categorical perception of violin articulation does not depend on the same auditory mechanism as is used for the phonetic distinction between affricate and fricative.

REFERENCE NOTE

1. Samuel, A. G. Personal communication, December 13, 1979.
2. Ganong, W. F., III. *An experiment on "phonetic adaptation"* (Progress Report, Vol. 116, pp. 206-210). Cambridge, Mass: Massachusetts Institute of Technology, Research Laboratory of Electronics, 1975.

REFERENCES

- ADES, A. E. Source assignment and feature extraction in speech. *Journal of Experimental Psychology: Human Perception and Performance*, 1977, **3**, 673-685.
- COLE, R. A., & COOPER, W. E. Perception of voicing in English affricates and fricatives. *Journal of the Acoustical Society of America*, 1975, **58**, 1280-1287.
- COOPER, F. S., LIBERMAN, A. M., HARRIS, K. S., & GRUBB, P. M. Some input-output relations observed in experiments on the perception of speech. *Proceedings of the Second International Congress of Cybernetics*. Namur, Belgium: Association International de Cybernetique, 1958.
- COOPER, F. S., & MATTINGLY, I. G. Computer controlled PCM system for investigation of dichotic speech perception. *Journal of the Acoustical Society of America*, 1969, **46**, 115(A).
- COOPER, W. E. Selective adaptation to speech. In F. Restle, R. M. Shiffrin, N. J. Castellan, H. R. Lindman, & D. B. Pisoni (Eds.), *Cognitive theory* (Vol. 1). Hillsdale, N.J.: Erlbaum, 1975.
- COOPER, W. E. *Speech perception and production*. Norwood, N.J.: Ablex, 1979.
- CUTTING, J. E., & ROSNER, B. S. Categories and boundaries in speech and music. *Perception & Psychophysics*, 1974, **16**, 564-570.
- CUTTING, J. E., & ROSNER, B. S. Discrimination functions predicted from categories in speech and music. *Perception & Psychophysics*, 1976, **20**, 87-88.
- CUTTING, J. E., ROSNER, B. S., & FOARD, C. F. Perceptual categories for musiclike sounds: Implications for theories of speech perception. *Quarterly Journal of Experimental Psychology*, 1976, **28**, 361-378.
- DIEHL, R. Feature analyzers for the phonetic dimension *stop* vs. *continuant*. *Perception & Psychophysics*, 1976, **19**, 267-272.
- DIEHL, R. L., ELMAN, J. L., & MCCUSKER, S. B. Contrast effects on stop consonant identification. *Journal of Experimental Psychology: Human Perception and Performance*, 1978, **4**, 599-609.
- EIMAS, P. D., & CORBIT, J. D. Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, 1973, **4**, 99-109.
- EIMAS, P. D., & MILLER, J. L. Effects of selective adaptation on the perception of speech and visual patterns: Evidence for feature detectors. In R. D. Walk & H. L. Pick, Jr. (Eds.), *Perception and experience*. New York: Plenum Press, 1978.
- GERSTMAN, L. J. *Perceptual dimensions for the frication portion of certain speech sounds*. Unpublished doctoral dissertation, New York University, 1957.
- LIBERMAN, A. M. Some characteristics of perception in the speech mode. In D. A. Hamburg (Ed.), *Perception and its disorders: Proceedings of the Association for Research on Nervous and Mental Disorders*. Baltimore: Williams & Wilkins, 1970.
- LIBERMAN, A. M., COOPER, F. S., SHANKWEILER, D. P., & STUDDERT-KENNEDY, M. Perception of the speech code. *Psychological Review*, 1967, **74**, 421-461.
- LIBERMAN, A. M., MATTINGLY, I. D., & TURVEY, M. T. Language codes and memory codes. In W. A. Melton & E. Martin (Eds.), *Coding processes in human memory*. Washington, D.C.: Winston, 1972.
- LISKER, L., & ABRAMSON, A. S. A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 1964, **20**, 384-422.
- LISKER, L., & ABRAMSON, A. S. Distinctive features and laryngeal control. *Language*, 1971, **47**, 767-785.
- LOCKE, S., & KELLAR, L. Categorical perception in a nonlinguistic mode. *Cortex*, 1973, **9**, 355-369.

- MACNEILAGE, P. F. Motor control of serial ordering of speech. *Psychological Review*, 1970, 77, 182-196.
- MILLER, J. D., WEIR, C. C., PASTORE, R. E., KELLY, W. J., & DOOLING, R. J. Discrimination and labelling of noise-buzz sequences with varying noise-lead times: An example of categorical perception. *Journal of the Acoustical Society of America*, 1976, 60, 410-417.
- PISONI, D. B. Identification and discrimination of the relative onset of two component tones: Implications for voicing perception in stops. *Journal of the Acoustical Society of America*, 1977, 61, 1352-1361.
- PISONI, D. B. Adaptation of the relative onset time of two-component tones. *Perception & Psychophysics*, in press.
- REMEZ, R. E. *An hypothesis of event-sensitivity in the perception of speech and bass violins*. Unpublished doctoral dissertation, University of Connecticut, 1978.
- REMEZ, R. E. Adaptation of the category boundary between speech and nonspeech: A case against feature detectors. *Cognitive Psychology*, 1979, 11, 38-57.
- REMEZ, R. E. Susceptibility of a stop consonant to adaptation on a speech-nonspeech continuum: Further evidence against feature detectors in speech perception. *Perception & Psychophysics*, 1980, 27, 17-23.
- SAMUEL, A. G., & NEWPORT, E. L. Adaptation of speech by nonspeech: Evidence for complex acoustic cue detectors. *Journal of Experimental Psychology: Human Perception and Performance*, 1979, 5, 563-578.
- SCHELLENG, J. C. The bowed string and the player. *Journal of the Acoustical Society of America*, 1973, 53, 26-41.
- SIMON, H. J., & STUDDERT-KENNEDY, M. Selective anchoring and adaptation of phonetic and nonphonetic continua. *Journal of the Acoustical Society of America*, 1978, 64, 1338-1357.
- STEVENS, K. N. The quantal nature of speech: Evidence from articulatory-acoustic data. In E. E. David & P. B. Denes (Eds.), *Human communication: A unified view*. New York: McGraw-Hill, 1972.

NOTES

1. This distinction is multiply cued. The difference between affricate and fricative may be realized synthetically by manipulating risetime or frication duration. In making our stimuli, we have covaried these two parameters as occurs in natural production of speech.

2. The proposal of Samuel and Newport may be untenable beyond this immediate experimental context; relevant counter-evidence already exists. Because periodicity-labeled risetime detectors at the very least preclude adaptation when adaptors are periodic and stimulus series are aperiodic (and vice versa), the findings of Cole and Cooper (1975) may be said to depart from the prediction in advance. In one study in their report, a [θa] adaptor (periodic, abrupt) caused identification boundaries to shift along a continuum ranging from [sa] (aperiodic, gradual) through [za] (periodic, gradual) to [θa]; the [θa] adaptor and [sa] test items could not have shared a detector tuned to periodicity and amplitude envelope, because the adaptor and test items differ in both attributes. Nevertheless, adaptation was observed. Ganong (Note 2) also describes an instance of adaptation that is inconsistent with the proposal of periodicity-labeled detectors. An aperiodic, gradual adaptor, [sæ], induced a shift in the phoneme boundary on a [bæ]-[dæ] continuum; this continuum consists of sounds that are both abrupt in amplitude rise and periodic. In the terms discussed by Samuel and Newport, this test series should not have been adaptable by aperiodic adaptors of any phonetic or auditory description. Both of these independent instances of adaptation count against the generality of the claims made by Samuel and Newport.

(Received for publication August 6, 1979;
revision accepted February 25, 1980.)