

Effect and artifact in the auditory discrimination of rise and decay time: Speech and nonspeech

MARCEL P. R. van den BROECKE
Utrecht University, Utrecht, The Netherlands

and

VINCENT J. van HEUVEN
Leyden University, Leiden, The Netherlands

The just noticeable difference (JND) for abruptness so far has not been studied in speech-like signals, and earlier studies have confounded decay time and overall signal duration. We therefore established JND for rise and decay time in a 10-80-msec range in a series of experiments involving an adjustment method using various speech and nonspeech signal types. Decay time did or did not vary independently of duration. Results showed that JND is in the order of 25%-30% of the reference value, with no essential difference between rise- and decay-time discrimination if these parameters are varied independently of duration. Decay-time discrimination turned out to be more accurate for wide-band signal types (noise) in the upper part of the stimulus range than it did for narrow-band signals (tones and complex harmonic signals). The data suggest that rise- and decay-time discrimination is too poor to reliably cue more than two categories in spite of the wide range of values found in speech sounds.

Linguistic Background

Differences in the degree of abruptness with which speech sounds begin or end signal phonemic or allophonic contrasts in many languages. The following instances of such linguistic contrasts for both vowels and consonants, as found in the literature, illustrate this point: (1) In the affricate-fricative distinction in English, /ʃ/ is differentiated from /ç/ by a relatively smooth vs. relatively abrupt onset of the noise burst (see Cutting & Rosner, 1974; Gerstman, 1957). (2) For Kabardian and Tlingit (languages spoken in the Caucasus and in Alaska, respectively), Jakobson, Fant, and Halle (1951, p. 23) claim that a pulmonic-ejective contrast is brought about by differences in the offset characteristics of fricatives, the ejectives having a more abrupt offset than the pulmonics.¹ (3) In French, vowels may or may not begin with a glottal stop. Malécot (1975) claims that differences in the onset time are an important acoustic and perceptual characteristic of this contrast. (4) Vowel offset differences have been claimed to correlate with short (checked) vs. long (free) vowels in Dutch (Cohen, Slis, & 't Hart, 1963).

Such contrasts have been accommodated in distinctive feature frameworks. Thus, Jakobson et al. (1951, p. 21) proposed two so-called envelope features. Of these, [continuant/interrupted] serves to separate sounds with relatively smooth onsets from those with relatively abrupt onsets of the amplitude envelope. The [free/checked] feature distinguishes segments with smooth vs. abrupt offsets. Later, Postal (1968, p. 71) adopted these features using different names: [+/- abrupt onset] and [+/- abrupt offset]. The importance of abruptness of amplitude change as an acoustic correlate of phonetic categories has recently been reaffirmed by Stevens (1980).

Acoustic Correlates

The envelope of a speech sound is acoustically defined by Jakobson et al. (1951) as the time function of the power of the speech waveform expressed in decibels integrated over a 20-msec time window. The perceived abruptness of onset is correlated with the rise time of the signal; similarly, the abruptness of offset has its main acoustic correlate in the decay time.

In the acoustic literature, decay time has been defined as the time needed for the signal to drop to 60 dB below its steady state intensity (i.e., an amplitude reduction to 1/1000). By the same token, rise time can be defined as the time needed for the signal to reach full intensity from a -60-dB reference point (Sabine, 1923). These definitions have been restated psychoacoustically in terms of the so-

We gratefully acknowledge valuable comments made by R. Plomp and S. G. Nootboom on an earlier version of this manuscript. M. P. R. van den Broecke's mailing address is: Institute of Phonetics, Utrecht University, Oudenoord 6, 3515 ER Utrecht, The Netherlands, V. J. van Heuven's address is: Leyden University, Schutteveld 9 (Room 606), 2316 XG Leiden, The Netherlands.

called "real" rise and decay time, in which the -60 -dB criterion was replaced by "the threshold of hearing" (Schuster & Waetzmann, 1929).

A definition of rise and decay time for speech signals should acknowledge that fact that speech does not contain steady state intensity portions as do the standard psychophysical signals. Accordingly, a measurement procedure for rise and decay time in speech signals has been proposed along slightly different lines: rise time is the time needed for the signal to increase from 10% to 90% of the peak intensity expressed in decibels; conversely, decay time is defined as the time needed for the signal intensity to drop from 90% to 10% (see Debrock, 1977).

Referring to the examples given above, the typical range of rise time in the affricate-fricative distinction is between 10 and 100 msec.² For example, the longest rise time found by Gerstman (1957) for fricatives pronounced in isolated words was 105 msec. Rise times for vowels are typically between 10 and 50 msec (Debrock, 1977; Malécot, 1975). Decay time of vowels is found to vary between 40 and 150 msec (Debrock, 1977; Lehto, 1969). To our knowledge, there are no published data on consonant decay times, but it seems safe to assume that these will normally not exceed the decay times found for vowels.

Perceptual Correlates

In spite of several convincing studies demonstrating the feasibility of separating signals into linguistically motivated categories on the basis of an acoustic difference in rise or decay time (see Gerstman, 1957; Kunisaki, Higuchi, & Fujisaki, 1978), it remains to be shown that these parameters are relevant from a perceptual point of view. There are no cases in which abruptness phenomena have been claimed to be the only physical correlate of a linguistic contrast. In all instances, concomitant acoustic changes in, for instance, overall segment duration, silent interval duration, spectral energy distribution, and rate of spectral change may be equally important. Perceptual experiments in which all concurrent parameters were controlled are conspicuously lacking. To our knowledge, only Gerstman (1957) has been (moderately) successful in this respect: it can be reconstructed from his data that a crossover from affricate to fricative is effected by increasing the rise time of the noise burst from 20 to 80 msec, but only if the overall duration of the noise burst is held constant at 100-120 msec (see van Heuven, 1979, Figure 2). For all other overall noise-burst durations, ranging, in 20-msec steps, from 40 to 180 msec, rise-time differences were inconsequential. Noise bursts lasting less than 100 msec were invariably perceived as affricates, and above 120 msec as fricatives.

Psychophysical Data on Rise/Decay Time JNDs

To our knowledge, no data are available on just

noticeable differences (JNDs) of rise- and decay-time differences for speech signals. However, von Békésy (1933) described a series of experiments with tone bursts, which we shall now review briefly.

JND for rise time was established by asking subjects to adjust the rise time of the second member of an identical pair of signals until a difference between the two members could be heard. Rise times were sampled in a 400-2,000-msec range, using 800-Hz tone bursts presented at 60 phon. The results are summarized in Figure 1.

Weber ratios were as small as 15% in the 1,500-2,000-msec part of the range, but increased steadily towards the lower end of the range, reaching values of 30% at the 400-msec sample point. Extrapolating from this trend into the speech range, one would expect the discrimination inaccuracy to increase even further.

In a second experiment, von Békésy measured JNDs for decay time. Here stimuli had 0-msec rise time and a 300-msec steady state duration. Results are again summarized in Figure 1. Decay-time Weber ratios were at or below 10% for reference values between 400 and 2,000 msec and tended to go up slightly (14%) only for the 300-msec-sample point.

In yet another experiment, von Békésy eliminated the 300-msec steady state portion, so that the stimuli decayed from the moment of onset onwards. In this condition, Weber ratios did not increase for shorter reference values, and were still in the 10% range for decay times of 100 msec (see Figure 1). Judging by the data of these three experiments, it might easily be concluded that decay times are systematically discriminated more accurately than are the corresponding rise times (10% JND for decay time vs. 15%-30% for rise times).

In our own previous experiment (van Heuven & van den Broecke, 1979), we established JNDs for rise and decay times of nonspeech signals in a manner similar to von Békésy's, but differing in the following two important respects: Our reference rise and decay times were sampled in a 0-100-msec range,

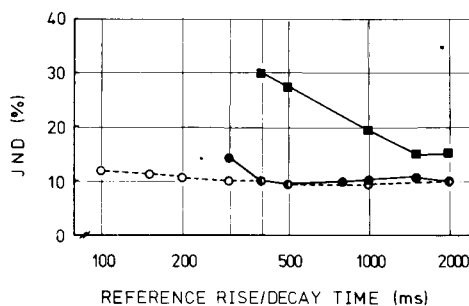


Figure 1. JND for abruptness phenomena ($\Delta T/T$ in %) for rise time (squares), decay time with 300-msec steady state portion (filled circles), and decay time with 0-msec steady state portion (open circles). After von Békésy (1933, Figures 7, 8, and 10; 1960, Figures 9-13, 9-14, and 9-16).

which largely covers the speech range for these phenomena, and the signal amplitude changed during the rise or decay portion as a linear function of time, whereas exponential rise and decay functions were employed by von Békésy. Linear functions show greater discontinuities in the amplitude envelope than do exponential functions, so that it is reasonable to expect that JNDs for linear functions will be smaller. Data from a related experiment by Miller (1948) may be interpreted as supporting this point of view: He found that 7-msec and 70-msec decay times were just noticeably different from each other for exponential functions, but that a 3.5-msec and a 35-msec decay-time difference was sufficient when linear functions were used.

No data are available for other decay time values. Rise times were not incorporated in this experiment.

Our own results, which will be discussed in some detail later, are compatible with von Békésy's with respect to rise time. Both for sine waves of 1000 Hz and for noise bursts, JND was about 25% for the major part of the stimulus range. However, the decay-time superiority effect encountered in von Békésy's results was not replicated in our experiment in the case of sine waves, for which discrimination accuracy was in fact slightly worse than it was for corresponding rise times. In the case of noise bursts, discrimination accuracy was significantly better for decay time than for the corresponding rise time, especially in the upper half of the range used, but the magnitude of this asymmetry did not even approach that found by von Békésy.

Potential Effect of a Double Cue

The electronic circuitry used to generate the stimuli for the experiments carried out both by von Békésy and ourselves was such that overall signal duration was constant for all rise-time values, but varied in the decay-time condition, since the decay-time portion was added to the steady state portion. Thus, in our own study (van Heuven & van den Broecke, 1979) the 0-100-msec decay-time range corresponded with a change in overall duration of 250-350 msec (Experiment 1) and 450-550 msec (Experiment 2). In von Békésy's experiments, decay times ranged from 300 to 2,000 msec, corresponding to a range in overall signal duration of 600-2,300 msec (Figure 1, filled circles) or of 100-2,000 msec, coinciding with the overall duration range (Figure 1, open circles).

Essential for the method of adjustment as a means of establishing JND is that reference and matching signals differ in one parameter only. In both von Békésy's experiments and our own, this was not the case. It is clear from the above description of the stimulus material that whenever two signals were unequal in terms of decay time, they also dif-

fered in overall duration. However, when these signals have relatively large steady state portions, the overall duration increment due to a longer decay time may be perceptually negligible so that the adjustment proceeds on the basis of the decay-time difference only.

A similar objection can be raised against the categorical perception experiments reported by Cutting and Rosner (1974), Diehl (1976), Kat and Samuel (1980), Remez, Cutting, and Studdert-Kennedy (1980), Rosen and Howell (1981), and Samuel and Newport (1979). In all these studies, rise time and overall signal duration covaried. For the purposes of these experiments, this procedure was justifiable, since the authors sought to create a continuum between two cognitive categories (pluck vs. bow, affricate vs. fricative, stop vs. continuant) by realistically covarying several parameters involved in the contrast. Such studies do not address the question of how to establish JNDs for each of the parameters at various places along the continuum used, and what their separate contributions to the contrast consist of.

Thus, in our earlier experiments (van Heuven & van den Broecke, 1979), the decay-time increment was 10 to 80 msec, corresponding to an 800% difference, which will be noticeable. The corresponding change in overall duration (steady state portion was 450 msec) from 460 to 530 msec is only 15%. We do not know how the 800% decay-time difference and the 15% duration difference compare in terms of possible perceptual dominance. In von Békésy's experiments, the possible weight of the duration cue is even greater: With a 0-msec steady state duration (Figure 1, open circles), any adjustment in decay time will be paralleled by an equal percentage change in overall duration, so that the relative weights of the two cues are equal. Thus, the shorter the steady state duration, the larger will be the cue value of the overall duration with respect to the cue value of the decay time.

In order to minimize the possibility of the occurrence of a double cue, the decay-time range should be small relative to the overall duration.

Obviously, this requirement is not met in von Békésy's experiments. In our previous study, a reduction in the overall duration range from 450-550 msec to 250-350 msec, increasing the potential influence of overall duration had no effect on JNDs for decay time, indicating that the influence of overall duration was roughly equal in both cases. When changing the overall duration to an even shorter value, one more realistic for speech signals (130-230 msec), however, the chance of overall duration's playing a significant role is increased. These considerations made it necessary to run further experiments, with duration either kept constant or

allowed to vary with decay time. Strictly speaking, this choice also leads to a double cue (MacMillan, Note 1), since, when total duration is kept constant, the duration of the steady state section is (inversely) correlated with the decay time. The ideal solution would be a full factorial design in which both duration and decay time were varied. Apart from the fact that this would lead to an unmanageably large number of signal conditions, this was not really the object of our experiment. Rather, we wanted to make rise-time and decay-time findings comparable, which they were not in von Bekesy's and our 1979 experiments. Even if a double cue remained in the present experiments—and our results show this to be highly unlikely—its effect would be the same for both rise- and decay-time conditions. We expected that the condition in which overall duration varied with decay time would lead to a greater accuracy of adjustment than is the case when overall duration is kept constant in spite of variations in decay time.

METHOD

Using either analog electronic switches with variable rise and decay times or a digital computer, sound bursts of various spectral compositions were given a variety of amplitude envelopes. As in our previous experiments (van Heuven & van den Broecke, 1979), the method of adjustment was used to estimate the threshold for reproduction of rise and decay time. The choice of this method was motivated largely by our wish to be able to compare our results with those of our previous experiments and of von Békésy.

Reference and comparison signals were presented in that order, with an interval of .5 sec and repeated every 4.2 sec. Subjects were asked to adjust a blind knob controlling the rise or decay time of the comparison signal (11 deg of rotation corresponding to a 1-msec change in rise/decay time) until they could no longer hear a difference between the reference signal and the comparison signal. The final setting was recorded with an accuracy of .1 msec, and then the next determination was initiated.

The starting value of the comparison signal alternated between the upper (110 msec) and the lower (15 μ sec) limits of the range employed, for each successive determination. The signals were presented binaurally at 60 dB above threshold, a level determined prior to each change in signal type, in a sound-treated booth (Amplifon GR 11) through headphones (Sennheiser HD 424). Rise and decay times of the reference signal were sampled at various points within the range from 0 to 100 msec. In a given session, either rise or decay time varied while the other ramp of the signal was given a constant duration of 50 msec. During the rise and decay portions of the signals, amplitude in volts changed as a linear function of time.

In each experiment, eight phoneticians, native speakers of Dutch, participated on a voluntary basis. The subjects were audiometrically normal and ranged in age between 22 and 38 years. They received no remuneration for their services.

In each experiment, the subjects made 128 threshold determinations, consisting of either four signal conditions with 16 sample points along the rise/decay time continuum (densely sampled) or eight signal conditions with 8 sample points (sparsely sampled), depending on the experiment (see descriptions of individual experiments). Each stimulus type occurred twice per block of stimuli (signal condition).

The order of presentation of the stimulus types within signal conditions, and of signal conditions within each experiment, was distributed over subjects according to a complete Latin square design so that possible order and learning effects were counterbalanced. The various signal conditions and stimulus values are listed below separately for each experiment. Table 1 presents a summary of the variables used. A more detailed description of each experiment is provided as follows:

Experiment 1

Sixteen sample points (0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, and 100 msec) were selected from either the rise or the decay time continuum. The signals used were synthesized Dutch low vowels, viz, /a/ and /a/ (OVE IIIb, see Liljencrants, 1968), with the following formant values (after Pols, Plomp, & Tromp, 1973):

	F1	F2	F3	F4	F5
/a/	800	1300	2600	3500	4000
/a/	680	1040	2600	3500	4000

All bandwidths were set at their respective midrange values, and fundamental frequency was held at a constant value of 130 Hz. When decay time varied, the steady state portion was kept constant at 130 msec, while overall signal duration covaried with decay time, thus giving a duration range of 50 (onset) + 130 (steady state) + 0 to 100 (decay) = 180-280 msec.

When rise time varied, the total duration was kept constant by compensating for longer rise times by shortening steady state duration. This seemingly irrational asymmetry between rise- and decay-time conditions is unavoidable when using analog gates as described by, for example, von Békésy (1933, 1960).

Signals were generated in real time and gated by means of electronic switches (Grason-Stadler 1287B) that had been modified so as to allow continuous rather than stepwise adjustment of rise or decay time. Control logic was provided by two programmable timers (Devices Digitimer D4030).

Experiment 2

This experiment was identical to Experiment 1 except for the source signals used. These were the synthesized (OVE IIIb) Dutch high vowels /i/ and /u/, with formant values:

	F1	F2	F3	F4	F5
/i/	295	2200	2700	3500	4000
/u/	340	810	2325	3500	4000

Experiment 3

This experiment was identical to Experiments 1 and 2, except that the signals consisted of sawtooth and triangular waves with a 130-Hz fundamental frequency (Krohn-Hite 5300 function generator) that had been digitally recorded (12 bits, 10 kHz, LP cutoff 4.5 kHz). Amplitude envelopes were generated by a computer in real time.

Experiment 4

In this experiment, the signals were identical to those of Experiment 3, except that the computer program was modified so as to maintain a constant overall duration, irrespective of the particular decay time used. To this end, longer decay times were compensated for by shorter steady state durations such that the total overall duration was kept constant at 230 msec. Rise time was kept constant at 50 msec.

Experiment 5

As stimuli for this experiment, the vowels /a/ and /a/ were used (see Experiment 1). Moreover, 1-kHz sine waves (Krohn-

Table 1
Summary of Experimental Conditions

Experiment	Rise Time	Steady Time	Decay Time	Total	Signal
1	0-100 (16 values) 50 fixed	180-80 fixed (16 values) 130 fixed	50 fixed 0-100 (16 values)	230 180-280 (16 values)	/a,α/
2	as	in	Experiment	1	/i,μ/
3	as	in	Experiment	1	sawtooth, triangle
4	50	180-80 (16 values)	0-100 (16 values)	230	sawtooth, triangle
5	10-80 (8 values) 50 fixed	170-100 (8 values)	50 fixed 10-80 (8 values)	230	sine wave, white noise, /a,α/

Note—All values are in milliseconds.

Hite 5300 function generator) and white-noise bursts (General Radio 1382 noise generator, LP cutoff 4.5 kHz) were used. As in Experiment 4, overall stimulus duration was kept constant at 230 msec throughout. Rise and decay times were sampled at eight points in 10-msec steps along a 10- to 80-msec continuum.

RESULTS

Table 2 summarizes the results of Experiments 1-5 in terms of the standard deviation (SD) of adjust-

ment for each reference value separated out for the experiments and conditions.

Figure 2 also summarizes the results of these experiments. SDs of adjustment are given for each of the signal conditions, averaged over the eight sample points shared by all signal conditions.

The upper panels of Figure 2 represent the results obtained in Experiments 1, 2, and 3, in which decay time covaried with overall duration. In the

Table 2
Summary of Experimental Results

	Stimulus Value																\bar{X}	α	β	r
	0	5	10	15	20	25	30	35	40	45	50	60	70	80	90	100				
Experiment 1																				
α Decay	4	6	7	6	6	7	8	7	9	11	11	12	10	11	8	7	8.8	5.4	.08	.85
a Decay	7	4	5	6	5	5	8	7	10	8	7	10	13	12	10	7	8.0	3.5	.11	.89
α Rise	1	4	5	5	6	11	10	15	13	14	20	19	17	22	26	13	13.1	3.2	.25	.92
a Rise	2	6	8	8	10	16	9	9	14	7	15	20	14	20	15	9	12.5	6.5	.15	.70
Experiment 2																				
i Rise	2	6	6	9	14	7	8	16	21	25	20	25	29	24	17	13	17.0	4.1	.32	.87
i Decay	3	4	5	10	12	11	11	12	16	14	16	15	8	20	14	15	12.5	8.0	.11	.61
u Rise	3	14	12	15	10	15	17	15	18	18	19	22	22	17	12	12	16.7	11.7	.12	.76
u Decay	4	4	15	6	6	6	10	7	13	11	14	10	14	12	15	15	10.3	7.7	.07	.43
Experiment 3																				
Sawtooth Rise	20	10	12	13	13	16	12	17	12	15	18	19	14	20	14	20	15.1	11.6	.09	.67
Sawtooth Decay	3	7	7	8	8	8	10	14	9	11	10	8	10	7	10	7	9.2	8.9	.01	.07
Triangle Rise	5	6	11	14	10	15	14	12	9	13	14	14	23	14	26	16	13.6	10.3	.08	.51
Triangle Decay	6	9	9	10	10	9	10	8	10	10	18	12	15	11	9	8	11.0	8.2	.07	.53
Experiment 4																				
Sawtooth Decay	2	12	10	9	9	14	16	17	16	16	19	22	19	27	22	17	16.2	7.1	.23	.93
Triangle Decay	1	5	9	6	11	12	10	19	14	19	21	21	22	25	18	18	15.8	5.4	.26	.92
Experiment 5																				
Sine-Wave Rise			7		9		14		16		17	15	15	14			13.4	8.9	.10	.69
Sine-Wave Decay			7		14		10		11		13	19	19	22			14.4	5.9	.19	.89
White-Noise Rise			6		8		8		10		14	14	19	18			12.1	3.5	.19	.97
White-Noise Decay			6		10		11		21		14	17	12	14			13.1	9.0	.09	.49
α Rise			9		7		14		14		13	19	19	16			13.9	7.3	.15	.84
α Decay			8		7		13		19		11	14	15	21			13.5	6.8	.15	.75
a Rise			6		7		9		10		11	15	17	17			11.5	3.7	.17	.98
a Decay			10		9		6		9		15	15	26	15			13.1	4.9	.18	.72

Note—Entries indicate SD of adjustment for each reference value for the various experiments and conditions, Experiments 1-5. \bar{X} represents the mean of the SDs; α indicates the intercept and β the slope of the linear regression line predicting SD as a function of the reference value. r is the correlation coefficient associated with the regression line. All statistics are based on the 10- to 80-msec part of the stimulus range.

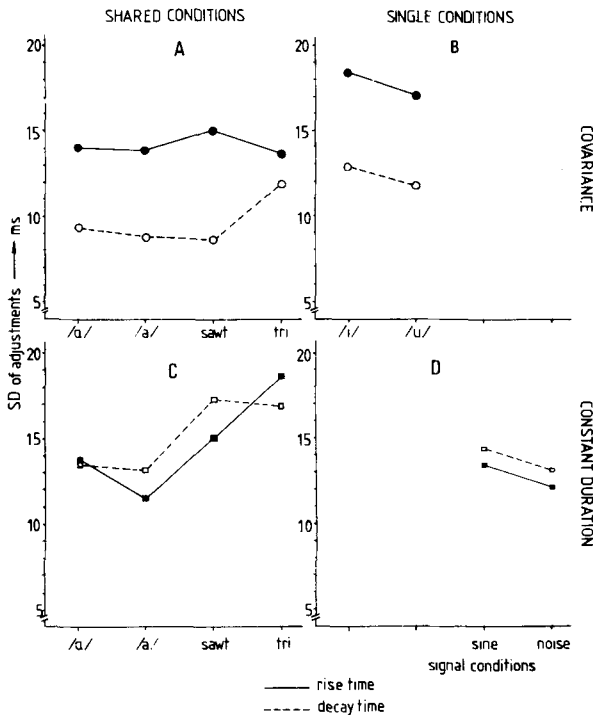


Figure 2. SD of adjustment per signal condition, averaged over eight stimulus values, with and without covariance of decay time and overall duration. "Shared conditions" refers to the signal types /a/, /a/, sawtooth, and triangle presented with and without covariance. "Single conditions" refers to signal types that were presented either with covariance (/i/, /u/) or without covariance (sine, noise). Solid lines represent rise time, and dotted lines decay time. The abscissa does not represent a continuous variable.

lower panels, results are given for Experiments 4 and 5, in which overall duration was kept constant. The results for the /a/, /a/, sawtooth, and triangle signal conditions are given in panels A and C (shared conditions). The results for vowels /i/ and /u/ are given in panel B and those for sine waves and noise bursts, in panel D (single conditions).

Next a selection of the data was made such that only the comparable signal conditions in the upper and lower panels of Figure 2 were analyzed further for JNDs. This selection is given in Figure 3, in which SDs of adjustment are plotted, pooled across signal conditions, for each of the stimulus values. SDs as a function of reference rise/decay time are given for rise- and decay-time values when decay time and overall duration covaried (filled and open circles, respectively) and when overall duration was kept constant (squares). Linear regression functions and the corresponding correlation coefficients are given in this figure for each of the four curves.

DISCUSSION

Interaction of Decay Time and Overall Duration

Figure 2 shows that, in Experiments 1 through 3

(covariance), SDs of adjustment are markedly smaller for decay time than for the corresponding rise-time condition. This decay-time superiority shows up in each of the six signal types used in the experiments. Experiments 4 and 5 (constant overall duration), however, reveal a slight decay time inferiority. The different behavior of rise and decay times in the two sets of experiments can be accounted for only by the presence of the double cue of overall duration and decay time in Experiments 1 through 3.

A classical three-way analysis of variance was performed on the shared signal conditions (see Figure 2) with signal type (/a/, /a/, sawtooth, triangle), presence vs. absence of covariance, and position of variable slope (rise vs. decay) as factors assuming fixed effects. The results indicate that presence vs. absence of covariance exerts a highly significant effect [$F(1,221) = 19.5, p < .001$]. Decay times have significantly smaller SDs than rise times [$F(1,221) = 13.5, p > .001$], but, as the significant interaction between these two factors indicates [$F(1,219) = 15.8, p < .001$], the decay-time superiority is restricted to the first three experiments, in which overall duration and decay times covaried.

In light of these results, our earlier conclusion that listeners are able to perceptually isolate offset duration from total stimulus duration turns out to be unwarranted.³ In the two experiments described in our earlier report (van Heuven & van den Broecke, 1979), in which we also used a 0-100-msec decay-

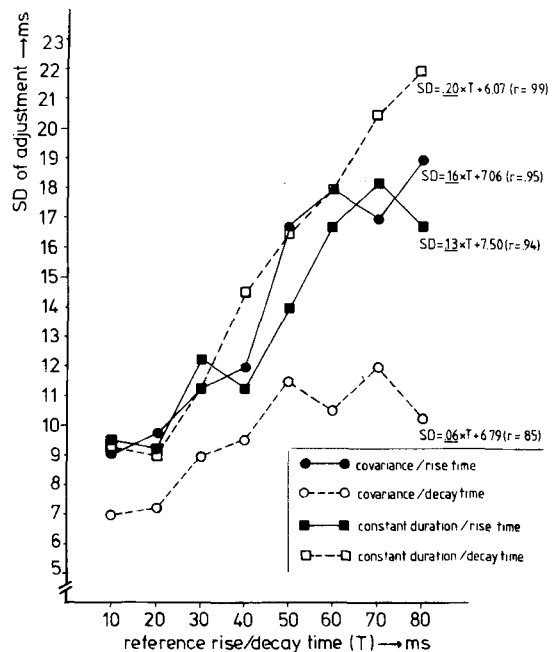


Figure 3. SD of adjustment as a function of reference rise/decay time in milliseconds, separated out for rise and decay times, with and without covariance of decay time and overall duration. Linear regression functions and correlation coefficients are indicated for each graph.

time range, the potential cue value of overall duration was much weaker, since we used steady state portions of 250 and 450 msec, respectively, as opposed to 180 msec in the present experiments. Ironically, reducing the steady state portion from 450 to 250 msec was largely inconsequential for decay-time accuracy, but a further reduction from 250 to 180 msec appears to have brought the overall duration cue within the noticeable range. The JND for decay time mentioned by von Békésy, 10%, is right in the middle of the 5%-15% JND range customarily found in the literature on overall duration discrimination, which lends further support to the correctness of our reinterpretation of his results.

Rise Time vs. Decay Time

On account of the overall duration artifact explained in the previous section, we shall exclude the decay-time results obtained in the first three experiments ("covariance") from further analysis. It will then become apparent that the discrimination accuracy for rise and decay times is essentially the same (see Figure 3). No significant differences in SD of adjustment between the remaining three conditions (i.e., constant duration/decay time, constant duration/rise time, covariance/rise time) could be established by a posteriori tests for contrasts (Newman-Keuls procedure, $p \leq .05$).

In view of these facts, the conclusion that rise- and decay-time discrimination is equally (in)accurate seems warranted.

Difference Limens

In accordance with Cardozo (1965) and Rakowski (1971), we adopted SD of adjustment as the measure for JND. Inspection of Figure 3 reveals that absolute JND increases from about 10 msec for the shortest reference values of T to about 20 msec for the longest values. Relative JNDs, expressed as Weber ratios $\Delta T/T$, decrease from around 100% for the 10-msec reference value to 25% for the 80-msec sample point.

Rather than giving an averaged JND for the entire rise- and decay-time range, as was done in the results section, one would like to be able to predict JND more precisely from the reference value of T by some simple function. As the statistics given in Figure 3 indicate, such predictions are quite adequately made by linear regression functions (correlation coefficients ranging between .94 and .99). The general regression function for the three curves under analysis is: $\Delta T = 7 + .16T$ ($r = .94$).

Auditory vs. Physical Decay

The nature of auditory decay. It may, at first, seem to be a remarkable coincidence that the classical psychophysical literature should contain no

data on rise- and decay-time limens within the range of speech. Yet, this seeming omission was clearly motivated, on the grounds of a postulated lower limit (or absolute threshold) for decay time, below which any offset would sound equally, that is, maximally, abrupt.

The existence of such an absolute threshold can be explained by assuming that neural activity does not end immediately after the cessation of the acoustic stimulus, but persists for some time until it is reduced to threshold level. Thus, any physical decay time shorter than the decay rate of neural activity will be masked.

This absolute threshold, called physiological decay time by von Békésy, turned out to be approximately 140 msec for 800-Hz tone bursts with exponential decay functions, and was essentially unaffected by the intensity of the physical stimulus. The nature of the decay rate of auditory sensation was explored in greater detail by, for example, Plomp (1964), who used a masking experiment in which probes of various intensities were presented at various intervals after the offset of the masker. His results show that, when expressed in decibels, the decay of poststimulus auditory sensation is a linear function of log time.

Linear vs. exponential decay functions. In this section, the terms "linear" and "exponential" refer to graphical representations in which the decay of signal intensity expressed in volts is plotted as a function of linear time.

As mentioned in the introduction, Miller (1948) found that the offset portion of noise bursts decaying to threshold as an exponential function of time had to exceed a critical duration of 70 msec in order to be perceptually distinguishable from (i.e., sound less abrupt than) an instantaneously switched-off signal. However, this critical duration turned out to be as short as 35 msec when a linear decay was used instead.

Given Plomp's (1964) description of the decay function of poststimulus auditory sensation, Miller's result seems to be understandable. To illustrate the point, we have replotted, in Figure 4, the auditory decay function as estimated by Plomp (thin line), as well as the 35-msec linear and the 70-msec exponential decays used in Miller's stimulus noise bursts. The decay of a physical stimulus will be auditorily indistinguishable from an instantaneous offset if, in terms of our figure, it remains below the (thin) line expressing poststimulatory sensation after an instantaneous offset (or rather, if the physical stimulus decay does not exceed the auditory decay by more than a critical amount).

Thus, a signal with a 10- or 20-msec offset, whether linear or exponential, will sound as abrupt as an instantaneously switched-off signal. However, for a

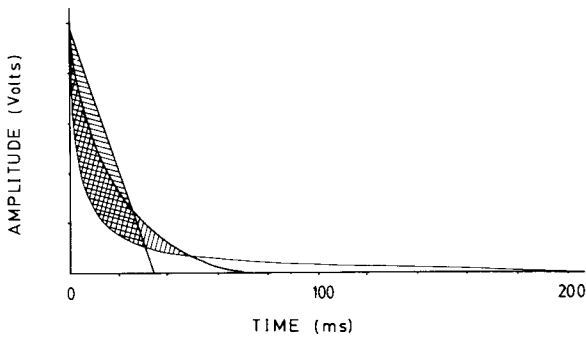


Figure 4. Decay of poststimulatory auditory sensation after an instantaneously switched-off white-noise burst (thin line), and the decay portions of two white-noise bursts, plotted along a linear amplitude scale (i.e., in volts) as functions of time (after Plomp, 1964). The 35-msec linear and 70-msec exponential decays should be just noticeably different from an instantaneous offset.

linear stimulus decay of 35 msec, the auditory decay line is exceeded sufficiently (cf. shaded area in Figure 4) to be perceived as more gradual than an instantaneous decay. Obviously, this perceptual effect will not obtain for an exponential decay of 35 msec, as this function still falls below the auditory decay, that is, is masked by it. Apparently, exponential signal decay functions must reach threshold no sooner than after 70 msec in order to be noticeably more gradual than an immediate offset.

Effects of spectral distribution. It has been demonstrated that the decay rate of auditory sensation differs for signals with various spectral characteristics. Miller (1948) replicated one of von Békésy's experiments using white noise instead of tones, and found that the absolute threshold for decay time ("critical time" in Miller's terminology) had decreased to about 70 msec, or to about half the critical time found for tonal stimuli.

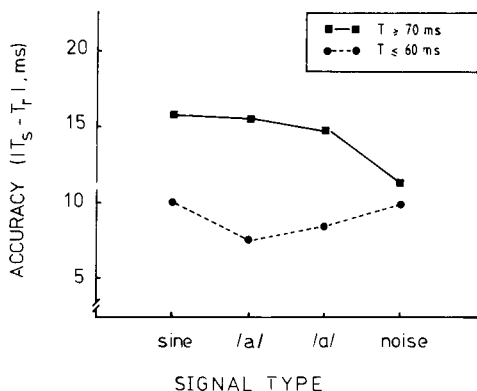


Figure 5. Accuracy of decay-time reproduction (absolute difference between stimulus and response value in milliseconds) plotted separately for the lower ($T \leq 60$ msec) and upper ($T \geq 70$ msec) parts of the stimulus range, for the four spectrally different signal types in Experiment 5. The ordinate does not represent a continuous variable.

Assuming that the values for the above absolute thresholds may indeed be halved when linear decay functions are used, we would predict very poor discrimination (strictly speaking, none at all) for linear decay times below 70 msec for sine waves or below 35 msec for noise bursts, or, in other words, better discrimination of decay time in noise bursts than in tones in the very restricted range of decay values between, say, 50 and 100 msec.

This prediction was clearly borne out by the results of our previous study (van Heuven & van den Broecke, 1979), in both of which experiments decay time discrimination was significantly more accurate for noise bursts than for tones in the upper half of the stimulus range, that is, for decay times between 50 and 100 msec.

In the present series of experiments, this type of effect will be more difficult to obtain since the crucial 90- and 100-msec sample points have been left out. By way of illustration, we have plotted, in Figure 5, mean accuracy-of-decay-time adjustment (defined as the absolute difference between stimulus and response; see van Heuven & van den Broecke, 1979) for each of the four signal types used in Experiment 5 (sine, white noise, /a/, and /a/), accumulated separately for the upper ($T \geq 70$ msec) and lower ($T \leq 60$ msec) parts of the stimulus range.

As expected, accuracy of adjustment is generally poorer in the upper part of the range [$F(1,510) = 28.0, p < .001$], according to a two-way analysis of variance with the dichotomized stimulus value parameter and signal type as factors. However, the eight means in Figure 5 turn out to be grouped such that only the sine and vowel signal types in the upper part of the range differ significantly from the other five conditions (Newman-Keuls procedure, $p \leq .05$ criterion), which do not differ from each other. Thus, only for white noise is decay time reproduction equally accurate in the lower and upper parts of the stimulus range.

REFERENCE NOTES

1. MacMillan, N. Personal communication, 1982.
2. van Heuven, V. J. J. P., & van den Broecke, M. P. R. *Auditory discrimination of rise and decay times in various speech and nonspeech sounds* (Progress Report of the Institute of Phonetics 5.1). Utrecht: Utrecht University, 1980.

REFERENCES

CARDOZO, B. L. Adjusting the method of adjustment: SD vs SL. *Journal of the Acoustical Society of America*, 1965, 37, 786-792.

CATFORD, J. C. *Fundamental problems in phonetics*. Edinburgh: Edinburgh University Press, 1977.

COHEN, A., SLIS, I. H., & 'T HART, J. Perceptual tolerances of isolated Dutch vowels. *Phonetica*, 1963, 9, 65-78.

CUTTING, J. E., & ROSNER, B. S. Categories and boundaries in speech and music. *Perception & Psychophysics*, 1974, 16, 564-570.

DEBROCK, M. An acoustic correlate of the force of articulation. *Journal of Phonetics*, 1977, 5, 61-80.

- DIEHL, R. Feature analyzers for the phonetic dimension stop vs. continuant. *Perception & Psychophysics*, 1976, **19**, 267-272.
- GERSTMAN, L. J. *Perceptual dimensions for the friction portions of certain speech sounds*. Unpublished PhD thesis, New York University, 1957.
- JAKOBSON, R., FANT, G., & HALLE, M. *Preliminaries to speech analysis: The distinctive features and their correlates*. Cambridge: MIT Press, 1951.
- KAT, D., & SAMUEL, A. G. More adaptation of speech by non-speech. *Journal of the Acoustical Society of America*, 1980, **68**, S10.
- KUNISAKI, O., HIGUCHI, N., & FUJUSAKI, H. Extraction of acoustic features and the classification of the voiceless affricates /ts/ and /ch/ in Japanese. *Journal of the Acoustical Society of America*, 1978, **64**, S179-180.
- LEHTO, L. *English stress and its modification by intonation: An analytic and synthetic study of acoustic parameters*. Helsinki: Suomalainen Tiedeakatemia, 1969.
- LILJENCRAFTS, J. The OVE III speech synthesizer. *IEEE Transactions on Audio and Electroacoustics*, 1968, **AU-16**, 137-140.
- MALÉCOT, A. The glottal stop in French. *Phonetica*, 1975, **31**, 51-63.
- MILLER, G. A. The perception of short bursts of noise. *Journal of the Acoustical Society of America*, 1948, **20**, 160-170.
- PLOMP, R. Rate of decay of auditory sensation. *Journal of the Acoustical Society of America*, 1964, **36**, 277-282.
- POLS, L. C. W., PLOMP, R., TROMP, H. Frequency analysis of Dutch vowels from 50 male speakers. *Journal of the Acoustical Society of America*, 1973, **53**, 1093-1101.
- POSTAL, P. *Aspects of phonological theory*. New York: Harper & Row, 1968.
- RAKOWSKI, A. Pitch discrimination at the threshold of hearing. In *Proceedings of the Seventh International Congress on Acoustics, Budapest*. Budapest: Akademiai Kiado, 1971.
- REMEZ, R., CUTTING, J., & STUDDERT-KENNEDY, M. Cross series adaptation using song and string. *Perception & Psychophysics*, 1980, **27**, 524-530.
- ROSEN, M., & HOWELL, P. Plucks and bows are not categorically perceived. *Perception & Psychophysics*, 1981, **30**, 156-168.
- SABINE, W. C. *Collected papers on acoustics*. Cambridge: Harvard University Press, 1923.
- SAMUEL, A., & NEWPORT, E. Adaptation of speech by non-speech: Evidence for complex acoustic cue detectors. *Journal of Experimental Psychology: Human Perception and Performance*, 1979, **5**, 563-578.
- SCHUSTER, K., & WAETZMANN, E. Über den Nachhall in geschlossenen Räumen. *Annalen der Physik*, 1929, **5**, 671-695.
- STEVENS, K. N. Acoustic correlates of some phonetic categories. *Journal of the Acoustical Society of America*, 1980, **68**, 836-842.
- VAN HEUVEN, V. J. The relative contribution of rise time, steady time, and overall duration of noise bursts to the affricate-fricative distinction in English: A re-analysis of old data. In J. J. Wolf & D. H. Klatt (Eds.), *ASA 50 speech communication papers*. New York: The Acoustical Society of America, 1979.
- VAN HEUVEN, V. J. J. P., & VAN DEN BROECKE, M. P. R. Perceptual discrimination of rise and decay times in tone and noise bursts. *Journal of the Acoustical Society of America*, 1979, **66**, 1308-1315.
- VON BÉKÉSY, G. Über die Hörsamkeit der Ein- und Ausschwingvorgänge mit Berücksichtigung der Raumakustik. *Annalen der Physik*, 1933, **16**, 844-860.
- VON BÉKÉSY, G. *Experiments in hearing*. New York: McGraw-Hill, 1960.

NOTES

1. For a different interpretation, however, cf. Catford (1977, p. 248, footnote 3).
2. Of these authors, only Debrock (1977) makes his measurement procedure explicit. It is not entirely clear to us how the results of the other studies mentioned here should be interpreted.
3. An earlier indication that decay-time and steady-time perception are not independent of each other was obtained in an unpublished experiment (van Heuven & van den Broecke, Note 2). In that study, a control condition was included in which the comparison stimulus was given a steady state duration that was twice as long as that of the reference stimulus (260 vs. 130 msec). Although accuracy of decay-time adjustment was essentially unaffected by this change, the results contained a remarkable effect: In the "duration mismatch" conditions, decay was reproduced at values some 20 msec shorter than they were for the "duration match" stimuli. Apparently our subjects were not able to suppress the need to overshoot the decay time in the "duration mismatch" comparison signals so as to approximate, or compensate for, the shorter overall duration of the reference stimulus.

(Manuscript received May 17, 1982;
revision accepted for publication December 15, 1982.)