948

The effects of breath sounds on the perception of synthetic speech

D. H. Whalen
Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511

Charles E. Hoequist
BNR, Inc., Box 13478, Research Triangle Park, North Carolina 27709

Sonya M. Sheffert
Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511 and Department of Psychology, University of Connecticut, Storrs, Connecticut 06269

(Received 8 March 1994; revised 23 November 1994; accepted 30 January 1995)

When preparing to speak, talkers typically take a breath. The perceptual effect of adding naturally produced breath intake sounds to synthetic speech was examined. In experiment 1, subjects were better at transcribing synthesized sentences that were preceded by a breath sound than those that were not, in addition to the improvement due to practice that is typically found with synthetic speech. Experiment 2 found that replacing the breath with the spectrally similar sound of rustling leaves had no effect on the accuracy. Experiment 3 had breaths before randomly selected sentences. Only the practice effect was significant, though there was a tendency for sentences with the breath sounds to be remembered better. In experiment 4, we tested whether the appropriateness of the breath sound to the sentence size (relatively short or long) affected the use of the breath sound. Appropriateness had no effect, perhaps because the range of sentence durations was too small. Experiment 5 replicated experiment 1 but used leaf sounds rather than silence in the nonbreath sentences. The presence of breath was again found to aid recall. Overall, the current results indicate that adding the breath intake sound to synthetic sentences improves listeners' ability to recall those sentences.

PACS numbers: 43.72.Ja, 43.70.Aj

INTRODUCTION

Speech is produced by the controlled shaping of the outgoing airstream, and so the lungs must have a certain amount of air in them before speech begins. There will be many times, then, when an utterance will be preceded by a breath intake. It is possible for speakers to make this intake relatively quietly, and singers, actors, and broadcasters are explicitly trained to try to avoid having their breath intake be audible. Yet it is also quite common for the breath intake to be apparent to the listener. The training for quiet breath intakes is not completely effective, as is easily shown by listening to a radio broadcast and focusing on the breaths. In everyday speech, therefore, it is more likely that the breath intake will be audible than not.

In the speech research literature, every aspect of the production of speech that has been found to affect the acoustic signal systematically has been shown to affect perception, for example, by changing the boundaries between similar segments (Studdert-Kennedy, 1976; Repp, 1982; Liberman and Mattingly, 1985; Lisker, 1986). It was our suspicion that the breath intake would as well. At the very least, the breath should indicate the imminent onset of speech. It is conceivable that the breath sound would give information about the speaker's vocal tract size as well, though this will not be tested here. Another possibility is that the duration of the breath intake would give information about the length of the upcoming utterance, since the duration of the breath intake seems to vary with utterance size (Atkinson, 1973; Hixon *et al.*, 1987; Whalen and Kinsella-Shaw, 1994) or syntactic structure and/or complexity (Grosjean and Collins, 1979; Conrad *et al.*, 1983; Sugito *et al.*, 1990). The present experiments represent a first look at the perceptual relevance of breath sounds.

I. EXPERIMENT

The first experiment was designed simply to see whether the breath sound affected perception at all. We compared the memory for sentences in the first half of a set of 40 sentences versus that in the second half. The breath sounds occurred either with the first 20 sentences or with the second 20 sentences.

A. Method

1. Stimuli

The sentences were modified from the stimuli developed by Lea (1974), which combined many words containing a particular sound or group of sounds (/s/, for example) within meaningful sentences. (The sentences are listed in the Appendix.) Most of the modifications were made to reduce the length of some of the sentences, in order to give us both short and long sentences. The short sentences averaged 8.1 words in length, and the long, 15.2 words. There were also six warm-up sentences that occurred at the beginning of each test. These were not scored.

The sentences were synthesized via Klattalk, an academic precursor to Dectalk (Klatt, 1972). Only three words had to be changed from standard orthography to phonetic

 0001-4966/95/97(5)/3147/7/$6.00

transcription in order for the synthesis to be correct. No further modifications were made, even though there were no doubt changes in the input strings that could have been made to make some of the sentences more comprehensible. We needed to have some errors to study, and the sentences were, in fact, understandable. Not only did the sentences sound correct to us, but every sentence was heard correctly by at least one of the subjects in the actual tests.

The breath sounds were recorded by the first author. His vocal tract is similar to that of the speaker whose voice served as the model for the voice of the synthesizer. The breath sounds were elicited by having the speaker prepare to read the next sentence. To ensure that a breath intake would occur, he exhaled before the cue to begin was given. The sentence was not actually read, so the breath was appropriate to the sentence only in the sense that the speaker was preparing for that sentence. From these 40 breaths, we selected three each for the short sentences and the long. The breaths before short sentences averaged 597.4 ms in duration, and those before long sentences, 738.4 ms in duration. We wanted to have the duration appropriate, so we kept the short breaths with the short sentences and the long breaths with the long sentences. Three different tokens were used so that any peculiarities of a particular breath sound would average out, and to prevent the sounds from becoming overly familiar to the listeners.

2. Procedure

There were four test orders, each containing the 40 test sentences plus the six practice items. Two sequences were used, one of which is shown in the Appendix. To balance any intrinsic difficulties in the sentences, half of the subjects received the sentences numbered 21–40 first, then those numbered 1–20. Two versions of each sequence were used. In one, breaths occurred before each of the sentences in the first half of the test and before none in the second half. In the other, breaths occurred before each of the sentences in the second half, but none before sentences in the first half. In this way, each sentence appeared equally often in the first and second half of the test, and equally often with and without a breath preceding it.

Each trial consisted of a warning tone, a breath (in the breath condition) and a sentence. The warning tone consisted of the first three harmonics of a 750-Hz fundamental, 600 ms in duration, with a linear amplitude ramp at both the beginning and the end. It was an average of 1.7 dB more intense than the peak amplitude of the breath, and 34.5 dB less than the typical first word of the sentence. The tone was followed by 500 ms of silence. The breath, if it occurred, was followed by 300 ms of silence. This value was chosen by the experimenters after listening to a variety of values. If there was no breath, then the sentence began at the point where the breath would have begun. We considered extending the time from the warning tone to sentence to match that of the breath sentences, but felt that this would not be equivalent. The breath sounds have a natural time course, and can be predicted to end at a certain point. The corresponding silences would simply be unpredictably long. The long sentences were followed by 25 s of silence, and the short, 15 s. Sequences were recorded onto DAT tape for later playback.

For the sequences with the breaths in the first half (the "breath-no breath" condition), the six warm-up sentences also had breaths before them. The sequences with breaths in the second half ("no breath-breath") had no breaths with the warm-ups.

Subjects were instructed to listen to these synthetic sentences and then write down as much of each as they could remember. They were told not to begin writing until the sentence ended, and they were to quit as soon as the warning tone for the next sentence was heard.

TABLE I. Effects of block (first half of the sentences versus the second half) and order (breath sounds occurring in the first or the second block), experiment 1.

	Order	Breath/no-breath	No-breath/breath
Block:			
	1	81.7	79.8
	2	83.7	85.1
Difference:		2.0	5.3

3. Subjects

The subjects were 32 colleagues from Haskins Laboratories and BNR, Inc. 14 were female, and 18 were male. Of these, 29 were native speakers of English, while three were fluent, non-native speakers of English. All volunteered their time. There were 16 in the breath-no breath order and 16 in the no breath-breath order.

B. Results

Subject responses were scored on a word basis. A correct word in the correct order counted as one. A word that was more than half correct (e.g., correct onset and vowel but incorrect final) counted as one half. A correct word in an incorrect location counted as one half. Function words and content words were counted equally. While this might have imposed an unfair burden on the listener for sentences from a human speaker due to the reduction of function words, the synthesizer did not mimic such reduction. Again, as mentioned before, every words was recognized by at least some of the listeners.

The overall percent correct values for each group are shown in Table I. In an analysis of variance with the within-factor of block (first half of the test versus second) and the between factors of breath (first or second block) and order (whether sentences 1–20 or 21–40 came first), there was no overall influence of breath or order [$F(1,28)<1$, n.s., for each]. Block was highly significant [$F(1,28)=33.34$, $p<0.001$], and the interaction of block with breath was significant [$F(1,28)=6.71$, $p<0.05$]. In separate analyses of the two breath groups, the no breath-breath group showed a significant improvement of 5.3% [$F(1,15)=50.33$, $p<0.001$], while the breath-no breath group showed a nonsignificant improvement of 2.0% [$F(1,15)=4.08$, $p<0.10$].

C. Discussion

The present results show both a common pattern, improvement with synthesis over time, and a new effect, improved recall with co-occurring breath intakes. This improvement over time has been found consistently (Pisoni and Hunnicutt, 1980; Greenspan *et al.*, 1988). Listeners seem to grow accustomed to the peculiarities of a particular synthetic voice, which reduces processing and improves their performance. In the present case, this effect was enhanced by the addition of breath sounds to the second half of the set of sentences. While removing the breath sounds in the second half did not make subjects worse in the second half, it did reduce the improvement to the point of statistical nonsignificance.

There are several possible explanations for this outcome. The breath sounds could have improved the naturalness of the synthetic speech by increasing the degree to which the utterances sound like they were produced by a human talker. Naturalness and intelligibility are highly correlated (Pavlovic *et al.*, 1990), but exactly how naturalness improves speech perception remains unspecified. Perhaps including aspects of natural speech events, such as the sound of a breath intake, might serve to place this rather unusual voice more in the range that the listener is familiar with. Resources that might otherwise be used in qualifying the voice as truly human can instead be used to extract more linguistic information from the signal.

Alternatively, the breath sounds may simply serve as an extra "warning tone," focusing subjects' attention on the task. If so, replacing the breath sounds with nonvocal tract sounds should confer similar benefits to intelligibility. This possibility is tested in experiment 2 by replacing the breath sounds with the sound of rustling leaves.

II. EXPERIMENT 2

In the first experiment, we deliberately varied the amount of time that occurred between the offset of the warning tone and the onset of the sentence depending on whether there was an intervening breath or not. We felt that leaving an equivalent amount of silence, while certainly simple to do, would not equate the two sets of sentences in terms of the time the subjects had to prepare. Specifically, the breath sounds would be equivalent to "filled silence," and the listener would be much better able to determine when that period would end, since the breath sounds have a natural offset. A better way to equate the two, it seemed, was to replace the breath sound with a sound that was equivalent in most ways, but irrelevant to the speech.

We chose to replace the breath sounds with the sound of rustling leaves, since the leaf sounds were naturally produced, very similar in amplitude contour, and similar in spectral composition. Both sounds are aperiodic and spread across the spectrum, though there is some speechlike shaping to the breath sounds. We matched the leaf sounds to the breaths in both duration and loudness, so that, as pure attention orienting sounds, they would be as equivalent as possible.

TABLE II. Effects of block (first half of the sentences versus the second half) and order (leaf sounds occurring in the first or the second block), experiment 2.

	Order	Leaf/no-leaf	No-leaf/leaf
Block:			
	1	83.6	81.7
	2	87.3	85.1
Difference:		3.7	3.4

A. Method

1. Stimuli

The sentences were those of experiment 1.

The leaf sounds were recorded in a sound-isolated booth with the same microphone that had been used for the breath sounds. The leaf sounds were recorded onto DAT tape (Panasonic SV-3700) and later input into the Haskins PCM system at 20 kHz without preemphasis (Whalen *et al.*, 1990). The sounds themselves were generated by jostling a small pile of leaves.

Six leaf sounds were selected to match the six breath sounds of experiment 1. They were matched in duration and in loudness, as judged by the first and third author. They were on average 5.1 dB more in peak amplitude than the breath sounds.

2. Procedure

The test was identical to that of experiment 1 except that the breath sounds were replaced by leaf sounds. Otherwise, the trials and the orders were the same.

3. Subjects

The subjects were 40 undergraduates at the University of Alaska who received course credit for their participation; 25 were female and 15 were male. There was an equal number of subjects in each order.

All were native speakers of English.

B. Results

Errors were scored as before. Overall percent correct for the two orders and the two conditions are shown in Table II. The improvement for the two groups was virtually identical. An analysis of variance with the within factor of block and the between factor of order ("leaf-no leaf" versus "no leaf-leaf") was performed. Block was significant [$F(1,38) = 20.33$, $p < 0.001$] but order was not [$F(1,38) = 1.01$, n.s.] The interaction was also not significant [$F(1,38) < 1$, n.s.].

C. Discussion

Replacing the breath sounds with leaf sounds eliminated the advantage of the presence of the extra sound. The sounds were, on average, somewhat more intense than the breath sounds (though they were of roughly equal loudness), yet memory for the items with the extra sound was not better than for those without. Thus the improved memory in the

first experiment does not seem to be due to extra attention given to the task that the added sound might have had encouraged. Instead, it looks as though the involvement with speech is crucial.

The results of experiments 1 and 2 suggest that the presence on breath sounds improve the intelligibility of synthetic speech. In the next two experiments, we examine whether the improvement due to the breath is specific to the sentence the breath sound occurs with or is, perhaps, one that affects the subject's overall reaction to the synthesis.

III. EXPERIMENT 3

In this experiment, rather than having all in the sentences in either the first half or second half of the list preceded by a breath sound, we placed the breaths in front of randomly selected sentences. If the benefit from the breath sound is only one of making the synthesizer sound more natural, perhaps an interspersing of breaths would be the most natural of all. If so, we would see no specific improvement for the sentences with the breaths. In contrast, if the breath sound improves just that sentence it occurs with, then there should be an effect for those sentences with the breath sound.

A. Method

1. Stimuli

The same order of the sentences from experiment 1 was used. Two variants of each of these two orders were created. For one variant, the presence or absence of the breath sound varied randomly from sentence to sentence, with the constraint that there be an equal number of breath and no breath sentences. The other variant was the complement of the first, with no breath where the first had breath and vice versa. This same order for the presence or absence of breaths was used for the other sentence order, so that the pattern was attached to different sentences. Since we still wanted to examine the learning effect, we analyzed the first and second half of the sentences separately. This led to one extra breath sentence and one fewer no breath sentence in some conditions, with the complementary pattern in the others. Short breaths were still used with short sentences, and long with long.

2. Procedure

The procedure was the same as for experiment 2.

3. Subjects

The subjects were 40 undergraduates from the University of Alaska or the University of Connecticut, who received course credit for their participation. The same stimulus tapes were used at both locations. All subjects were native speakers of English. 24 were female and 16 were male. An equal number received each test order.

B. Results

Errors were scored as before. The overall pattern is shown in Table III. Note that in this case, both condition and presence of breath are within factors, since each subject had some of each. Breath is not a significant factor [$F(1,39)=1.96$, n.s.]. Block is a significant factor [$F(1,39)=13.03$, $p<0.001$]. Although there is the appearance of greater improvement for the sentences without breath sounds, the interaction of breath and block is not significant [$F(1,39)<1$, n.s.].

TABLE III. Effects of block (first half of the sentences versus the second half) and presence or absence of the breath sounds, experiment 3.

	Sentence condition	Breath	No-breath
Block:			
	1	84.5	81.2
	2	84.9	83.2
Difference:		0.4	2.0

C. Discussion

From these results, it appears that the improvement due to the breath sounds is a general one, not one specific to the sentences with the breath sounds. Although the sentences with the breaths were recalled more accurately (by 2%), this difference was not reliable. Unlike experiment 1, this experiment had both measures within subject, which would normally lead to a more sensitive measure. However, this also meant that half as many sentences were involved in each subject's scores, and the effects of individual sentences seems to have added noise. Perhaps a study with more sentences would have shown an effect. On the other hand, it may simply be that the effect is, indeed, a general one of improving the perceived naturalness of the synthesis. After all, the sentences were separated by at least 15 s of silence, and when speakers have a long time to take a breath, they can do so more slowly and therefore more quietly. If the listeners were construing the test as a fluent discourse, though (by ignoring the pauses during which they were writing), they may have assumed the speaker did not need a breath before each sentence anyway, since they are all relatively short. Two or three of them could easily be said by a human speaker without taking a breath. So the alternation of breaths and lack of breaths may have seemed appropriate to the usual pattern in which speakers do not take a breath before every sentence (Grosjean and Collins, 1979). (Of course, most speakers do not include a warning tone before each sentence!) The added confidence that the listener may have had, based on the presence of the breath sound, that this was a human speaker may have needed occasional reinforcement. It may not have been necessary to have a breath with each sentence to obtain the improvement, leading to the pattern we found in this experiment.

In addition, the breaths that were used were always ones that had been produced in anticipation of a sentence of a particular size. The next experiment examines whether anticipated sentence length, to the extent suggested by the breath duration, was also playing a role.

IV. EXPERIMENT 4

In the current experiment, we examine whether the two classes of breath sounds, those produced for the longer sen-

TABLE IV. Effects of block (first half of the sentences versus the second half) and appropriateness of the breath (to long or short sentences), experiment 4.

		Appropriate	Inappropriate
Block:			
	1	80.9	80.1
	2	84.3	84.6
Difference:		3.4	4.5

tences versus those produced for the shorter, would affect memory for synthetic sentences differently if put with the other type of sentence. Speakers seem to take a larger breath before longer sentences (Atkinson, 1973; Hixon *et al.*, 1987; Whalen and Kinsella-Shaw, 1994), so the information provided by the duration of the breath intake may be useful when listening to the resulting speech. This experiment tests this notion by presenting subjects with both appropriate breaths and inappropriate ones, to see if one is more helpful than the other or, indeed, whether the inappropriate breaths actually degrade performance.

A. Method

1. Stimuli

The sentences from the previous experiments were used. With the same randomizing scheme as was used for the alternating breaths in experiment 3, we presented subjects with both appropriate and inappropriate breaths before the sentences. That is, a long breath before a short sentence was deemed inappropriate, and similarly with a short breath before a long sentence. The sentences with breath in experiment 3 were identical (the "appropriate" sentences), while the sentences without breath in experiment 3 had the "inappropriate" breaths.

2. Procedure

The sequencing scheme of experiment 3 was used, with the inappropriate breaths occurring where the sentences with no breath appeared in that experiment.

3. Subjects

The subjects were 40 University of Connecticut undergraduates who received course credit for their participation. All were native speakers of English. 26 were female and 14 were male. There were equal numbers for each sequence.

B. Results

The overall results are shown in Table IV. An analysis of variance was performed with the within-subject factors of condition, appropriateness, and duration (short versus long sentences). Condition was again a significant factor [$F(1,39)=30.03$, $p<0.001$], as was duration [$F(1,39)=89.97$, $p<0.001$]. Appropriateness was not a significant main effect [$F(1,39)<1$, n.s.], and did not enter into any significant interactions, although there was one marginal interaction, with duration [$F(1,39)=3.70$, $p<0.10$]. Regardless of appropriateness, long breaths led to better performance than the short breaths. The short sentences averaged 88.9% for the short (appropriate) and 90.0 for the long (inappropriate), while the long sentences averaged 74.7% for the short (inappropriate) and 76.3% for the long (appropriate). The interaction is not significant, but is perhaps one worth pursuing in future studies. It may be that the longer breaths are better indicators of vocal tract size, or set up the expectation that a longer sentence will be produced (which would presumably leave excess processing capacity for the short sentences).

C. Discussion

Appropriateness of the breath intake did not influence recall in this experiment. This may indicate that there is no specific information to be gained from the breath intake. But there are several reasons why an effect of appropriateness might not have appeared in this study. The sentences were short (even the long ones were only 17 words), and so there might not have been enough duration difference to influence the breath intake. An analysis of the data of Whalen and Kinsella-Shaw (1994) reinforces the notion that the breath duration might not be a strong indicator here. If we select the sentences from their corpus with 7–10 or 14–18 syllables (our short and long sentences, respectively), we find 21 short and 27 long sentences. There is a small difference in the duration of the breath intakes for those sentences in the expected direction (421 ms vs 461), but this difference is not significant. With isolated utterances, such as those used here, there is too much flexibility on the part of the talker about whether to take a breath for the duration to be highly predictive. In a long discourse, there must be breaks for breath intake (if the speaker is human), and those are probably more constrained. So not only must we avoid making too much of the null hypothesis, we should also be aware that there may be more revealing circumstances to be had.

V. EXPERIMENT 5

The last experiment allows for a more direct comparison of conditions from experiments 1 and 2, using a within-subjects design in which one set of listeners is exposed to both speech and nonspeech precursors. In experiment 1, we found a significant effect of breath on the recall of synthesized sentences, while experiment 2 demonstrated no improvement from nonspeech leaf sounds. At the very least, this suggests that the inclusion of some introductory sound resulted in an improvement, and that the effect is strongest when the sound originates from a vocal tract. In order to confirm that breath sounds can reliably facilitate the recall of synthesized speech, we changed the experimental design. The effects of both breath and leaf sounds were evaluated using a within-subjects design in which the same set of listeners received both kind of precursor sounds.

A. Method

1. Stimuli

The test order and sentences from experiment 1 and 2 were used. There were two test orders created, which presented sentences 1–20 first, or 21–40. Two versions of each

TABLE V. Effects of block (first half of the sentences versus the second half) and order (breath sounds occurring in the first or the second block, or leaf sounds occurring in the first or the second block), experiment 5.

	Order	Breath/leaf	Leaf/breath
Block:			
	1	80.7	79.4
	2	82.7	83.9
Difference:		2.0	4.5

order were created which either had a breath sound before each sentence in the first half, followed by leaf sounds before each sentence in the second half, or, conversely, the leaf sentences preceding the breath sentences.

2. Procedure

The procedure was identical to the previous experiments.

3. Subjects

The subjects were 74 undergraduates from the University of Alaska or the University of Connecticut, who received course credit for their participation. The same stimulus tapes were used in both locations. In order to place more confidence in these results, we approximately doubled the number of subjects we tested compared with experiment 1. All were native speakers of English. 39 were female and 35 were male. There were 37 in each order (breath-leaf or leaf-breath).

B. Results

Errors were tabulated in the same manner as the previous experiments. The overall percent correct values for each group are shown in Table V. As in the previous experiments, there is a practice effect, with performance improving in the second block. In addition, the means suggest that the recall accuracy across blocks was magnified when the breath followed the leaf sounds, relative to the leaf following the breath sounds. An analysis of variance was performed with the between factors of breath (present in the first block and followed by the leaf sounds, or vice versa) and the within factor of block (first half of the test versus second). There was no influence overall of breath or order [$F(1,72)<1$, n.s.]. Block was highly significant [$F(1,72)=32.16$, $p<0.001$], and the interaction of block with breath was significant [$F(1,72)=4.88$, $p<0.05$], confirming the recall advantage for the Leaf/Breath test order.

In separate analyses of the two breath groups, the breath-leaf group showed a significant improvement of 2.1% [$F(1,36)=4.61$, $p<0.05$], and the leaf-breath group also showed a highly significant improvement of 4.5% [$F(1,36)=44.41$, $p<0.0001$].

C. Discussion

This experiment confirms the finding of experiment 1, that breath sounds are helpful in recalling synthetic sentences. Even when directly contrasted with the acoustically similar leaf sounds, there is more improvement in the second block of the experiment when the breath sounds occur in that block. While this experiment does not allow us to speculate any more on the mechanism responsible for the increased retention, it does establish that an introductory aperiodic sound originating from a vocal tract is critical to the effect.

VI. GENERAL DISCUSSION AND CONCLUSION

In experiments 1 and 5, we found that placing the sound of a breath intake before synthesized sentences improved subjects' recall of those sentences. This result was not due to mere attention-orienting, since similarly constructed sounds of rustling leaves had no effect on recall. The effect may be due to an overall improvement in the perceived acceptability of the synthetic voice, since the improvement did not reach significance when the breath sentences were randomly mixed with the no breath ones. Even though the size of the breath intake may be a cue for the size of the upcoming sentence, the present set of stimuli was too narrow to tell whether the subjects were using this information or not.

As text-to-speech systems continue to improve, new sources of refinement must be sought. Terken and Lemeer (1988), for example, found that when the segmental quality was poor, the quality of the intonation did not matter. But when the segmental structure was clear, improving the intonation increased the overall acceptability ratings. Since the effects of the breath intake are relatively small, it is only with modern, generally successful synthesis that including them would be worth the effort. The size of the effect found here, around 1%–1.5%, may seem small, but any small effect can look large if the initial error rate is itself small. If we had a system with an overall error rate of 7.5%, for example, then a 1.5% improvement (to 6%) would represent a 20% reduction in the total error rate, a substantial contribution. For synthesis systems, the ideal way to include breath sounds would be to synthesize them. This may take some work, since the signals are very low in amplitude and composed of aperiodic noise. Experiment 2 showed that not any aperiodic noise will count as breath intake, so enough of the features of breath noise will have to be included for any benefit to accrue. This may actually be harder than synthesizing the speech signal itself, because the breath signal is so weak. For the moment, a hybrid system, which uses a set of digitized natural breath sounds, might be the most effective.

Fully effective use of breath noise, however, will likely be tied in with improvements in discourse management. Breath intake occurs mainly at syntactic boundaries (Grosjean and Collins, 1979; Conrad *et al.*, 1983; Sugito *et al.*, 1990), and its characteristics are probably under the same type of control as pausing, changes in pitch range, etc. Getting the breath right, then, will be part of the larger challenge of getting the discourse right. For the moment, however, there may be some gains made simply by filling an occasional pause with breath noise.

ACKNOWLEDGMENTS

The research reported here was supported by NIH Grant HD-01994 to Haskins Laboratories. We thank Carol A. Fowler, Jeff M. Kinsella-Shaw, and three anonymous review-

ers for helpful comments. Portions of this research were presented at the Spring (Ottawa) and Fall (Denver) meetings of the Acoustical Society of America, 1993.

APPENDIX

List of test sentences used in all experiments.

1. I doubt if he is the type that would like to exploit the flight.
2. Lloyd is proud to guide the town.
3. Sue was sick of setting tables and serving salad and suppers.
4. See if you can soak Sadie's socks.
5. In a close shave, the ship shook and shuddered from the shell, but showed no damage.
6. Puss is making a mess of the fresh fish.
7. He can slurp up soup in a snap.
8. We will sweep and mop the step, then wipe up the soap so you don't slip.
9. Tom takes tap water to make tea.
10. Your mutt sat on my suit with a lot of dirt on his feet.
11. According to the book, the baby should burp on the bib and be ready for a bath.
12. Both Bud and his boss beat me to the booth.
13. Ted heard from Bud that the food is good and the service is bad.
14. The code will aid the bid for the deed.
15. She refused to face the fact that her folks had to feed the fish and the fox.
16. Fern made a fuss over her foot.
17. The thief was safe in the cliff until his cough and his laugh gave the oaf away.
18. He couldn't hear the surf from the bluff.
19. They always rev up their engines, shove them in gear, and give a wave as they leave.
20. She had the nerve to move the stove.
21. The judge said Joe was the jerk who broke jail and stole my jacket and jeans.
22. Jeff will sell John that jeep next June.
23. Their mood was meek as they met to make the most of their merger.
24. The mob will miss the map the mutt took.
25. Before we began to sing, he had sung the song that you sang last year.
26. Ben says that soon the plan will be drawn.
27. I will read the rules about the role of the rook and the risks it can take.
28. He was rash to race after the wreck.
29. You should work the wood, wait a week, and then wash and wax it.
30. I wish you would woo and wed.
31. The troop will trudge down the trail on this part of its trek to the fort.
32. The crew craned their necks to see the crook.
33. The queen had a quirk of being on a quest for every quote that was quaint.
34. You should quit going to the quack.
35. Her dream was weird, in that her dress was drab and she drifted about like a drunk.
36. He told us to bring bread to the brunch.
37. They boast that his fist is fast and the worst to overcome.
38. My taste for the East is lost.
39. Under the strain, the strap stretched and straw was strewn all over the street.
40. It struck me that he strove to be strict.

Atkinson, J. E. (1973). "Aspects of intonation in speech: Implications from an experimental study of fundamental frequency," Ph.D., University of Connecticut, Storrs.

Conrad, B., Thalacker, S., and Schönle, P. (1983). "Speech respiration as an indicator of integrative contextual processing," Folia Phon. **35**, 220–225.

Greenspan, S. L., Nusbaum, H. C., and Pisoni, D. B. (1988). "Perceptual learning of synthetic speech produced by rule," J. Exp. Psychol.: Learn., Mem., Cog. **14**, 421–433.

Grosjean, F., and Collins, M. (1979). "Breathing, pausing and reading," Phonetica **36**, 98–114.

Hixon, T. J., Watson, P. J., and Maher, M. Z. (1987). "Respiratory kinematics in classical (Shakespearean) actors," in *Respiratory Function in Speech and Song*, edited by T. J. Hixon (College-Hill, San Diego, CA), pp. 375–400.

Klatt, D. H. (1982). "The Klattalk text-to-speech system," in ICASSP-82 (IEEE, New York), pp. 1589–1592.

Lea, W. A. (1974). "Sentences for testing acoustic phonetic components of systems," Sperry Univac, Report No. PX 10952.

Liberman, A. M., and Mattingly, I. G. (1985). "The motor theory of speech perception revised," Cognition **21**, 1–36.

Lisker, L. (1986). "Voicing in English: A catalogue of acoustic features signalling /b/ versus /p/ in trochees," Lang. Speech **29**, 3–11.

Pavlovic, C. V., Rossi, M., and Espesser, R. (1990). "Use of the magnitude estimation technique for assessing the performance of text-to-speech synthesis systems," J. Acoust. Soc. Am. **87**, 373–382.

Pisoni, D. B., and Hunnicutt, S. (1980). "Perceptual evaluation of MI-Talk: The MIT unrestricted text-to-speech system," in ICASSP-80 (IEEE, New York), pp. 572–575.

Repp. B. H. (1982). "Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception," Psychol. Bull. **92**, 81–110.

Studdert-Kennedy, M. (1976). "Speech perception," in *Contemporary Issues in Experimental Phonetics*, edited by N. J. Lass (Academic, New York), pp. 243–293.

Sugito, M., Ohyama, G., and Hirose, H. (1990). "A preliminary study on pauses and breaths in reading speech materials," Annu. Bull. RILP **24**, 121–130.

Terken, J., and Lemeer, G. (1988). "Effects of segmental quality and intonation on quality judgments for texts and utterances," J. Phon. **16**, 453–457.

Whalen, D. H., and Kinsella-Shaw, J. M. (1994). "Exploring the relationship of breath intake to utterance duration," J. Acoust. Soc. Am. **96**, 3327(A).

Whalen, D. H., Wiley, E. R., Rubin, P. E., and Cooper, F. S. (1990). "The Haskins Laboratories' pulse code modulation (PCM) system," Behav. Res. Methods Inst. Comp. **22**, 550–559.