NRL Report 8589

UNCLASSIFIED

The Perception of Pitch Waver in Synthetic Vowels Heard over Headphones and Loudspeakers

ASTRID SCHMIDT-NIELSEN AND STEPHANIE S. EVERETT

Communication Systems Engineering Branch
Information Technology Division

May 28, 1982

NAVAL RESEARCH LABORATORY
Washington, D.C.

Approved for public release; distribution unlimited.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NRL Report 8589	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) THE PERCEPTION OF PITCH WAVER IN SYNTHETIC VOWELS HEARD OVER HEADPHONES AND LOUDSPEAKERS		5. TYPE OF REPORT & PERIOD COVERED Final report on one aspect of an NRL problem. 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Astrid Schmidt-Nielsen and Stephanie S. Everett		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Research Laboratory Washington, DC 20375		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS PE 61153N; Proj. RR0210542; NRL Problem 75-0129-0
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Arlington, VA 22217		12. REPORT DATE May 28, 1982 13. NUMBER OF PAGES 14
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Voice naturalness
Perceptual tests
Pitch waver

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The perception of pitch waver in synthetic vowels was investigated. The waver was more easily detected when heard over a loudspeaker than over headphones. Measurements indicated that in a live-room environment there were amplitude fluctuations in the vowels, in addition to the intended pitch modulation, and that the combined modulation was better detected than the pitch changes alone. Waver was detected better for vowels than for pure tones.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

UNCLASSIFIED

CONTENTS

INTRODUCTION 1

EXPERIMENT 1 1

Method 2
Stimuli 2
Subjects and Procedure 2
Results and Discussion 2

EXPERIMENT 2 4

Method 4
Results and Discussion 4

EXPERIMENT 3 6

Method 7
Results and Discussion 7

CONCLUSIONS 10

ACKNOWLEDGMENTS 11

REFERENCES 12

UNCLASSIFIED

THE PERCEPTION OF PITCH WAVER IN SYNTHETIC VOWELS HEARD OVER HEADPHONES AND LOUDSPEAKERS

INTRODUCTION

When a person attempts to sustain a vowel sound at a constant pitch, there is usually a certain amount of waver or fluctuation in the pitch. (Pitch waver as used here refers to relatively gradual changes in fundamental frequency, whereas pitch jitter would refer to period-to-period variations.) This may be due to imperfect muscular control, and the extent of the waver is affected by various factors such as age and disease. Such fluctuations in pitch or fundamental frequency probably occur during ordinary conversational speech as well, but other normal pitch variations (both the trends due to overall prosodic influences and local variations due to inherent phoneme differences) would make these fluctuations difficult to observe except in unusually quavery voices. Unlike normal human voices, synthesized speech can be produced with a perfectly constant pitch, and any desired variations over time must be specified. Adding small amounts of pitch waver to synthesized speech should serve to make the synthetic voice sound more natural and human and less machinelike. Synthesized vowel utterances serve as a useful basis for investigating the perception of voice-pitch waver.

The perception of pitch differences in synthesized vowels has been investigated by Flanagan and Saslow [1]. They measured difference limens for fundamental frequency by having subjects listen to pairs of vowels. It has been repeatedly observed that the minimum detectable change ΔF for pure tones is different for pulsed sines than for frequency-modulated tones (e.g., by Fastl [2], among others). Voice-pitch waver is similar to frequency modulation, whereas the Flanagan and Saslow data are more comparable to research using pulsed tones. Unlike pure tones, speech sounds are complex and rich in harmonics. When the fundamental frequency of a vowel is modulated, the change in the harmonics is proportional to the number of the harmonic, so the actual frequency change will be considerably greater than the frequency change for the fundamental. Flanagan and Saslow noted that, while listeners probably make use of frequency information in the harmonics, the utilization is less than might be predicted from pure-tone data. A similar lack of direct comparability can be expected with regard to the perception of frequency-modulated tones and voice-pitch waver in vowels.

Loudspeakers were considered by Zwicker and Gässler [3] to be unsuitable for studying the perception of frequency-modulated tones because room resonances would alter the signal. Headphones offer more precise control over what the listener is receiving. In real situations, people hear speech, both natural and synthesized, in open room environments as well as over headphones or handsets. In order to test the perception of waver both in a sound field and with the signal coming directly to the ear, the present tests were carried out using a loudspeaker as well as headphones.

EXPERIMENT 1

The first experiment was designed to determine the discriminability of different modulation frequencies (f_ms). The values of f_m covered by this experiment can be informally described as ranging across three categories. At about 2 to 4 Hz, the modulation sounds like normal voice waver. This agrees with measurements taken on normal nonpathological voices [4]. At about 7 to 10 Hz, the modulation is best described as having a vibratolike character, and around 12 to 15 Hz the vowel takes on a distinctly buzzy sound which no longer seems very human in character. At high modulation frequencies, the fluctuations are no longer heard separately, but as sidebands. This aspect is of more interest for the study of tones than for speech phenomena, since it is well outside the range of normally produced voice fluctuations.

Manuscript submitted February 19, 1982.

Method

Stimuli

The synthetic vowel /a/ was used for the waver comparisons. The vowels were generated by an analog serial resonance synthesizer using the Peterson and Barney [5] formant values for /a/. The mean fundamental frequency, f_0, was 99 Hz. The waver around the mean f_0 value was generated using a low-frequency sine-wave oscillator as input to the f_0 control function of the synthesizer. The frequency excursion Δf (measured from highest to lowest frequency) and the modulation frequency f_m could be varied separately by varying the amplitude and frequency of the oscillator respectively. Vowel pairs (Same or Different) were recorded on a Nagra IV-S tape recorder, whose wow and flutter were measured to be less than one-third of the lowest waver generated in these experiments and, therefore, could be considered negligible. Each /a/ lasted 1.25 s, with 0.62 s between the vowels in the pair and 4.36 s between pairs. A trapezoidal function was used to shape the vowel onset and offset to give a natural-sounding stimulus.

In this experiment, the unmodulated vowel and 18 frequency-modulated vowels with values of f_m from 2.10 to 15.4 Hz were used. In the course of generating the stimuli, it became apparent that two vowels with the same f_m sounded alike only if the ups and downs of the modulation were alike with respect to the onset and offset of each vowel in the pair. This constraint influenced the selection of the f_m values that were actually used: 0, 2.10, 2.85, 3.18, 3.75, 4.28, 4.80, 5.30, 5.85, 6.40, 7.50, 8.50, 9.00, 10.2, 11.2, 12.3, and 15.4 Hz. In spite of the unequal intervals between stimuli, it was possible to make comparisons between pairs differing in f_m by approximately 0.5 Hz, approximately 1 Hz, and approximately 2 Hz. Five tapes were constructed using different random orders of each of 54 Same or Different pairs. The Δf for the modulated vowels was always 1.75 Hz in this study.

Subjects and Procedure

Subjects were 25 University of Maryland students who volunteered to participate for extra course credit. All subjects listened to all five tapes, yielding five trials on each pair. In addition to judging whether each pair was Same or Different, they rated their confidence in the judgment using 0, 1, or 2 to indicate no confidence to high confidence. Even though testing a few highly trained subjects tends to give more stable psychophysical data than testing a larger number of relatively naive subjects, we chose the latter course as being more representative of the way ordinary people hear speech sounds. Two listening conditions were used to compare the everyday situation of listening in a room environment with the laboratory situation of listening over headphones. Twelve subjects heard the stimuli over a Realistic 8-inch loudspeaker with built-in amplifier, and the remaining 13 heard the same stimuli over Koss PRO/4AA headphones.

Results and Discussion

Confidence ratings were used to compute A_g for each subject [6]. The measure A_g is derived from signal-detection theory and represents the area under the ROC (receiver operating characteristic). The different confidence levels were used to represent different criterion values. The measure A_g is equivalent to the proportion correct in a forced-choice task [7], and the detectability measure gives a control for response bias.

Table 1 shows the performance for various contrasts for the headphones and the loudspeaker. For both conditions, there was a general trend toward poorer discrimination as f_m increased. This trend was clearer and more pronounced for the loudspeaker data than for the headphone data. In general, subjects who heard the stimuli over headphones performed at a near-chance level, whereas subjects hearing the same stimuli over the loudspeaker were able to discriminate quite well. Table 2 shows the headphones-loudspeaker difference based on the average A_g for each subject.

UNCLASSIFIED

Table 1 — Performance of Subjects Making Same-Different Judgments of Vowel Pairs Differing in Modulation Frequency (Chance performance is 0.50)

Stimulus Pair (Hz)	f_m Difference (Hz)	Mean A_g	
		Headphones	Loudspeaker
3.18-3.75	0.57	0.520	0.714
3.75-4.28	0.53	0.583	0.654
4.28-4.8	0.52	0.613	0.688
4.8-5.3	0.50	0.575	0.672
5.3-5.85	0.55	0.599	0.661
5.85-6.4	0.55	0.578	0.484
8.5-9.0	0.50	0.498	0.538
2.10-3.18	1.08	0.551	0.886
2.85-3.75	0.90	0.541	0.791
3.18-4.28	1.10	0.646	0.813
3.75-4.8	1.05	0.725	0.826
4.28-5.3	1.02	0.661	0.802
4.8-5.85	1.05	0.552	0.710
5.3-6.4	1.10	0.615	0.714
6.4-7.5	0.90	0.512	0.661
7.5-8.5	1.00	0.535	0.615
9.0-10.2	1.20	0.577	0.559
10.2-11.2	1.00	0.582	0.670
11.2-12.3	1.10	0.521	0.488
12.3-13.4	1.10	0.444	0.495
0-2.10	2.10	0.654	0.961
6.4-8.5	2.10	0.522	0.829
9.0-11.2	2.20	0.603	0.683
10.2-12.3	2.10	0.459	0.670
11.2-13.4	2.20	0.515	0.603
13.4-15.4	2.20	0.518	0.613

Table 2 — Comparison of Vowels Heard over a Loudspeaker or over Headphones (Scores are based on average A_g for each subject; chance performance is 0.50)

Source	Mean A_g	Standard Deviation	t
Loudspeaker	0.70	0.108	4.21*
Headphones	0.56	0.042	

*$p < 0.001$

Both the listening medium and the modulation frequency influenced the subjects' ability to detect the modulation and to discriminate differences in modulation frequency. For the contrast between the unmodulated vowel and $f_m = 2.10$ Hz, the mean A_g was 0.96 for the loudspeaker group and 0.65 for the headphones group. For all of the contrasts between different f_ms, the mean loudspeaker A_gs were higher than the comparable A_gs for headphones. The overall averages were 0.70 for the loudspeaker and 0.56 for headphones. There was also a general trend indicating poorer discrimination of f_m differences with increasing f_m. A difference in f_m of 0.5 Hz gave an A_g of $0.68 - 0.71$ for the loudspeaker and $0.58 - 0.61$ for headphones at f_m values of 3 to 5 Hz, but both groups showed chance performance (near 0.50) at f_ms of 8 to 9 Hz. A 1-Hz f_m difference gave $A_g = 0.79 - 0.89$ for the loudspeaker and $0.55 - 0.65$ for headphones at $f_m = 3$ to 5 Hz, and A_g fell to near chance at $f_m = 11$ to 13 Hz. Both groups showed the same trend, and it is likely that the overall poorer discrimination for the headphones group was the result of the fact that the waver was less detectable over the headphones. The frequency response of the loudspeaker and headphones was measured, and it did not show any substantial differences that might account for the difference in waver detectability.

EXPERIMENT 2

A series of single-listener tests was conducted to eliminate several possible explanations for the loudspeaker-headphones difference. There are a number of differences between listening over headphones and listening over a loudspeaker that could conceivably result in the signal being perceived differently, but many of these may be irrelevant to the detection of voice waver. The effect of room resonances in altering the signal that reaches the ear is clearly a strong possibility. Other factors that might be considered include effects due to binaural differences and signal intensity.

Method

A new series of 10 randomized tapes was generated, as described above, in which Δf was systematically varied. The comparison pairs consisted of six Same pairs and six Different pairs. The Same pairs had Δfs of 0, 0.35, 0.44, 0.70, 1.11, and 1.75 Hz. The Different pairs consisted of the following contrasts: 0/0.35, 0/0.44, 0/0.70, 0/1.11, 0/1.75, and 0.70/1.75. These 12 pairs were produced at each of three f_ms, 2.1, 4.3, and 9.0 Hz, for a total of 36 test pairs on each of the 10 tapes.

One listener, CM, was tested under a variety of listening conditions. The equipment and the Same-Different judgments with confidence ratings were the same as in Experiment 1. Five tapes were used for each listening condition; and since comparable results were obtained for the three f_ms, this gave a total of 15 responses for each Δf comparison. The listener was told to listen for the waver as the basis for making the judgments but was otherwise naive as to the purpose of the experiment or any hypotheses guiding the choice of listening conditions. Listening levels were adjusted to be "comfortable and audible." The listener was seated about 1 m from the loudspeaker, at which point the signal was approximately 74 dB sound pressure level (SPL). The signal over the headphones was normally about 72 dB SPL.

Results and Discussion

The various listening conditions and their results are summarized in Table 3. Performance was again measured in terms of A_g. Values are given for each contrast, with Same pairs serving as controls (false alarms) for each contrast. Consider first the top two lines in Table 3, in which loudspeaker and headphone performances are contrasted. Since 0.50 indicates chance performance, the waver is probably somewhat detectable if the score exceeds 0.65; 1.00 indicates perfect performance, and we may wish to consider a score of 0.75 as the detection threshold. Even the smallest Δf was detected over the loudspeaker, while the largest did not exceed threshold over the headphones. It is clear that the waver was heard much better over the loudspeaker than over the headphones.

UNCLASSIFIED

Table 3 — Performance (A_g) for listener CM making Same-Different Judgments under Various Listening Conditions (Chance performance is 0.50)

Listening Condition	A_g for Contrast Pair Δf						
	0/0.35	0/0.44	0/0.70	0/1.11	0/1.75	Mean for 5 Pairs	0.70/1.75
Loudspeaker	0.91	0.95	0.92	0.97	0.98	0.95	0.47
Headphones	0.45	0.52	0.53	0.67	0.70	0.57	0.76
Headphones off the head	0.62	0.70	0.88	0.92	0.96	0.82	0.57
Loudspeaker low volume	0.59	0.76	0.91	0.98	1.00	0.85	0.58
Headphones low volume	0.53	0.56	0.55	0.72	0.88	0.65	0.76
Headphones with air leak	0.50	0.47	0.51	0.69	0.83	0.60	0.67
Loudspeaker anechoic chamber							
Facing speaker	0.48	0.47	0.55	0.83	0.82	0.63	0.50
Side to speaker	0.52	0.47	0.70	0.77	0.74	0.64	0.50
Back to speaker	0.45	0.59	0.43	0.74	0.76	0.59	0.53
Headphones recorded from mikes							
Stereo	0.51	0.39	0.62	0.74	0.81	0.61	0.55
Mono (to both ears)	0.43	0.50	0.56	0.72	0.79	0.60	0.55
Loudspeaker (one ear only)	0.40	0.71	0.84	0.81	0.85	0.71	0.53
Headphones (one ear only)	0.48	0.46	0.45	0.52	0.74	0.53	0.60
Headphones from mikes in live room							
Stereo	0.82	0.95	0.92	0.98	0.98	0.93	0.53
Mono	0.78	0.75	0.95	0.90	0.95	0.87	0.47

For the 0.70/1.75 contrast, both members were modulated, the subject had to discriminate differences in the *extent* of waver, and discrimination was better with headphones. Over headphones the Δf of 0.70 was not detectable, but to the degree that the Δf of 1.75 was, the two would be perceived as different. Over the loudspeaker, both stimuli would sound wavery, but the difference in degree might not be detected. Loudspeaker listening improves waver detection, but it does not necessarily enhance discrimination. Therefore, a high performance on this pair suggests that the threshold was actually above 0.70, whereas poorer performance on this pair suggests a lower threshold.

As a test of headphone response, the headphones, instead of being worn, were placed on the table directly in front of the listener so that they acted effectively like a loudspeaker but at a lower volume (about 62 dB SPL). This resulted in detection performance almost as good as that for the loudspeaker and well above normal headphone performance. This confirmed our observation, based on oscilloscope readings, that the headphones did respond to small pitch changes (as would indeed be expected of hi-fi headphones).

Zwicker [8,9] determined modulation thresholds for frequency-modulated pure tones and found that a smaller Δf was needed to exceed threshold as loudness increased. (Headphones were used to collect these data.) For comparison purposes the loudspeaker and headphone tests were repeated at low volumes (about 62 dB SPL for the loudspeaker; less than 60 dB SPL for the headphones). If we consider the original loudspeaker test and compare it to the two lower-volume tests—loudspeaker low and headphones off—it does indeed seem that the threshold was lower (i.e., scores higher) at the greater loudness. However, the normal-volume and low-volume headphone data do not agree with such a tendency. The headphone threshold in both cases was considerably higher (scores lower) than for any of

the loudspeaker conditions, and the low-volume condition, surprisingly, seems to have improved performance slightly. Intensity may be an important factor in the perception of voice waver, but intensity differences clearly do not account for the observed difference in detectability over loudspeakers and over headphones. It would not be surprising if frequency-modulation detection for speech sounds differs from that for pure tones, but this is a separate issue.

The possibility that the enclosed space over the ear may have affected perception was tested, even though the stimuli were well above threshold and low-frequency physiological masking should not have been a problem. The headphones were propped 2 to 3 cm away from the head using foam pads, and as can be seen this manipulation did not improve performance.

The room in which the tests were made was quite live, and the possibility that room reverberations contributed to the perception of the waver was considered. Three tests were conducted in an anechoic chamber: the listener on axis facing the loudspeaker, at 90° to the loudspeaker (one ear toward and one ear away from the loudspeaker), and at 180° with his back to the loudspeaker. The results for all three orientations are quite similar to those for the headphones and differ markedly from those for the loudspeaker in the live room. This suggests that room acoustics play an important role in the phenomenon. It is not clear, from this demonstration alone, whether room characteristics enhance the pitch waver in the signal or whether the perceptual system makes use of the added resonances in a more complex manner. It can be added that loudspeaker tests in at least two other live rooms also showed good performance (low thresholds), so the effect was not due to the particular characteristics of any one room.

In loudspeaker listening, binaural differences can play a role, whereas over headphones the signal to the two ears is identical. To assess the influence of binaural effects a stereo recording was made in a sound booth using two ALTEC Model 650 BL cardiod dynamic microphones placed about 20 cm apart, with one microphone 10 to 15 cm further from the loudspeaker than the other, with the result that one channel was slightly delayed and attenuated relative to the other. A comparable monophonic (mono) recording used the output of one microphone for both channels. These tapes were then played over headphones. As can be seen in the table, the results in both cases were comparable to other headphone data. One-ear performance was also assessed for the loudspeaker (with an ear plug and an ear defender) and for the headphones. This resulted in slightly lower performance in both cases, but the large loudspeaker-headphone difference was maintained. When the microphone tapes were remade in an acoustically live room instead of in a sound booth, headphone performance on these tapes was as good as with the loudspeaker in the live room, and there was no difference between the stereo and mono recordings. It appears that the loudspeaker effect depends more on room acoustics than on binaural differences.

A comparison of the three f_ms used in this experiment indicated that, on the whole, the waver was more detectable at the modulation frequency of 4.3 Hz than at 2.1 or 9.0 Hz. This agrees with the data of Zwicker [8] for pure tones, in which there was a slight lowering of threshold at a modulation frequency of about 4 Hz. It is interesting to note that this is also about the waver frequency one would expect from measurements of normal human voice waver [4]. The enhancement of waver detection in the presence of room acoustics, however, occurs at all modulation frequencies.

EXPERIMENT 3

The third experiment was conducted to verify and extend the findings of the previous experiment. The single-subject results were suggestive and eliminated a number of possibilities while strongly implicating the influence of room resonances. The most suggestive listening conditions for Experiment 2 were repeated with multiple subjects, new listening conditions were added, and acoustic measurements of the original recording and a live-room recording were made.

UNCLASSIFIED

Although the detection of vowel modulation has some aspects in common with modulation detection for pure tones, it is not suprising to find that there are differences as well. As was pointed out earlier, the harmonic structure of vowels carries additional information about the waver besides the changes in the fundamental frequency. The third experiment was conducted to investigate the contribution of high- and low-frequency information in the vowels to waver detection. The modulated vowels were high-pass filtered and low-pass filtered, and the filtered versions were tested. For comparison purposes, frequency modulation of a 1000-Hz tone and a 250-Hz tone was also tested.

Method

New sets of stimulus tapes were made for each of five test groups: unfiltered /a/, high-pass filtered /a/, low-pass filtered /a/, 1000-Hz sine tone, and 250-Hz sine tone. The vowels were generated as described above. For the filtered vowels, the synthesizer output was passed through an Allison 2 AB and 2 SKL filters. For both high-pass and low-pass filtering, the filters were adjusted to 0 dB response at 400 Hz, for high-pass to −40 dB at 200 Hz, and for low-pass to −40 dB at 600 Hz. The range of values of Δf to be used for each stimulus set was selected by informal listening. The step size between successive Δfs was either 2 dB or 4 dB, depending on the range to be covered.

Fifteen series of randomized Same-Different pairs were generated for each of the five stimulus sets and recorded as above. For each stimulus set, there were three listening conditions. The loudspeaker and headphones were as previously described. Live-room recordings (for headphone listening) were made by playing the tapes on the Nagra IV-S tape recorder over the Realistic loudspeaker and recording the output with two Altec 650 BL dynamic microphones, placed about 19 cm apart and about 1 m from the loudspeaker, and a Sony tape recorder.

Subjects were recruited and tested as in Experiment 1. Each subject was tested under all three listening conditions (loudspeaker, headphones, and live-room tape over headphones) for one of the five stimulus sets. The number of subjects tested was 15 for the unfiltered vowel, 10 for the low-passed vowel, 12 for the high-passed vowel, 10 for the 250-Hz tone, and 11 for the 1000-Hz tone. They heard five series of Same-Different pairs for each listening condition, and the A_g for each Δf was computed on the basis of each subject's confidence ratings.

The original recording and the live-room recording of waver samples for the unfiltered and filtered vowels and for the 250-Hz tone were evaluated to determine the exact frequency fluctuation and amplitude fluctuation, if any, of the stimuli. The waver samples were processed through a device developed for larynx pathology studies, which measures vocal periods and amplitude accurately on a period-by-period basis. The resulting information is displayed on a running basis of equivalent frequency and amplitude of successive periods, which may then be evaluated in terms of variation in fundamental frequency (in hertz) and variation in amplitude (in decibels). The amplitude measured is the highest peak amplitude within each period. A more complete description of the device may be found in Ref. 10. This device could not be used for the 1000-Hz tone. The intended modulation values for this series were checked by measuring the sample with the most extreme modulation on a narrowband spectrogram made on a Kay Sonograph model 7030A using a 20- to 2000-Hz frequency range.

Results and Discussion

Table 4 shows the performance of the listeners for each set of stimuli. The Δf for the stimuli is shown above each set of results. The values of Δf for all vowel stimuli refer to the extent to which f_0 was modulated. For the tone stimuli, Δf is the actual frequency excursion. The measurements were made after the subjects had been tested. In general the extent of waver was selected so as to be clearly perceptible in the extreme stimuli, but for the high-passed vowels the actual Δfs that were used were somewhat lower than was desirable.

Table 4 — Average A_g for Same-Different Judgments under Various Listening Conditions (Chance performance is 0.50)

Listening Condition	Number of Subjects n	A_g for Contrast Pair Δf										Δf at Which A_g Exceeds	
		0/0.35	0/0.44	0/0.55	0/0.70	0/0.88	0/1.10	0/1.40	0/1.80	0/2.20	0/2.80	0.75	0.60
Unfiltered /a/	15												
Loudspeaker		0.54	0.69	0.75	0.76	0.81	0.81	0.85	0.91	0.91	0.92	0.70	0.44
Live-room tape		0.46	0.52	0.53	0.65	0.77	0.82	0.94	0.94	0.88	0.94	0.88	0.70
Headphones		0.54	0.50	0.51	0.52	0.59	0.59	0.68	0.72	0.77	0.82	2.20	1.40
		0/0.11	0/0.18	0/0.28	0/0.44	0/0.70	0/1.10	0/1.80	0/2.80	0/4.40	0/7.00		
Low-pass filtered	10												
Loudspeaker		0.52	0.47	0.58	0.58	0.54	0.73	0.78	0.90	0.92	0.92	1.80	1.10
Live-room tape		0.50	0.48	0.41	0.46	0.50	0.57	0.45	0.73	0.75	0.85	4.40	2.80
Headphones		0.48	0.50	0.46	0.46	0.47	0.51	0.58	0.64	0.73	0.85	7.00	2.80
		0/0.10	0/0.13	0/0.16	0/0.20	0/0.25	0/0.30	0/0.50	0/0.60	0/0.80	0/1.00		
High-pass filtered	12												
Loudspeaker		0.49	0.49	0.53	0.63	0.62	0.60	0.64	0.69	0.78	0.76	0.80	0.20
Live-room tape		0.53	0.50	0.52	0.49	0.58	0.53	0.56	0.59	0.72	0.69	—	0.80
Headphones		0.48	0.50	0.51	0.55	0.48	0.52	0.52	0.52	0.46	0.55	—	—
		0/1.20	0/1.80	0/2.90	0/3.70	0/4.60	0/5.80	0/7.30	0/9.20	0/11.60	0/18.30		
250-Hz tone	10												
Loudspeaker		0.52	0.53	0.63	0.71	0.88	0.90	0.89	0.93	0.96	0.97	4.60	2.90
Live-room tape		0.51	0.40	0.58	0.58	0.60	0.77	0.80	0.85	0.90	0.95	5.80	4.60
Headphones		0.43	0.47	0.48	0.50	0.55	0.54	0.68	0.86	0.92	0.98	9.20	7.30
		0/0.30	0/0.50	0/0.80	0/1.30	0/2	0/3	0/5	0/8	0/13	0/20		
1000-Hz tone	11												
Loudspeaker		0.50	0.52	0.48	0.53	0.56	0.48	0.65	0.82	0.87	0.91	8	5
Live-room tape		0.49	0.51	0.57	0.55	0.53	0.55	0.70	0.80	0.90	0.94	8	5
Headphones		0.52	0.51	0.57	0.52	0.52	0.59	0.57	0.71	0.86	0.90	13	8

On the whole, the results were similar to those of the preceding experiments. For all five stimulus sets detection performance was better for the loudspeaker than for headphones, repeating and extending the results of the previous experiments. For the unfiltered vowels, live-room tape performance was almost as good as loudspeaker performance. For the other stimulus sets the live-room tape led to better detection than headphones, but not as good as the loudspeaker. Live-room tape performance was closer to loudspeaker performance for the high-passed vowels and the 1000-Hz tone than for the low-passed vowels and the 250-Hz tone. On the whole, the waver detection was best for the harmonically rich intact vowels, and these also showed the greatest loudspeaker-headphones difference. Detection was also better for the filtered vowels than for the pure tones, but the difference was not as great as might be expected if the subjects were able to make full use of the information carried in the vowel harmonics.

The similarities and differences in performance on the five stimulus sets, when taken in the context of the frequency and amplitude measurements, suggest some possible explanations of the detection differences, but they also present some contradictions. The values of Δf given in Table 4 represent the results of the frequency measurements (changes in f_0 for all vowel stimuli and actual frequency excursion for the tones). The frequency modulation was identical for all of the original tapes and the corresponding live-room recordings, but the amplitude measurements (see Fig. 1 and Table 5) showed some suggestive differences. For the original stimuli, the unfiltered vowels and the high-passed vowels had some amplitude modulation in addition to the frequency modulation. This apparently results from the manner in which pitch is generated by the synthesizer. The low-pass filtered stimuli and the 250-Hz tone, in contrast, had no amplitude changes in the original stimuli. The unfiltered vowels showed a

UNCLASSIFIED

Table 5 — Amplitude Fluctuation for Selected Stimuli

Test Condition	Δf	Original Fluctuation (dB)	Live-Room Fluctuation (dB)
Unfiltered /a/	2.8	~ 0.7	3
	1.75	~ 0.5	2
	0.88	*	1
Low-pass filtered /a/	7.0	*	3
	2.8	*	2
	0.7	*	1
High-pass filtered /a/	1.0	0.2-0.3	*
	0.8		
250-Hz tone	9.2	*	2
	3.7	*	1

*Negligible (see Fig. 1)

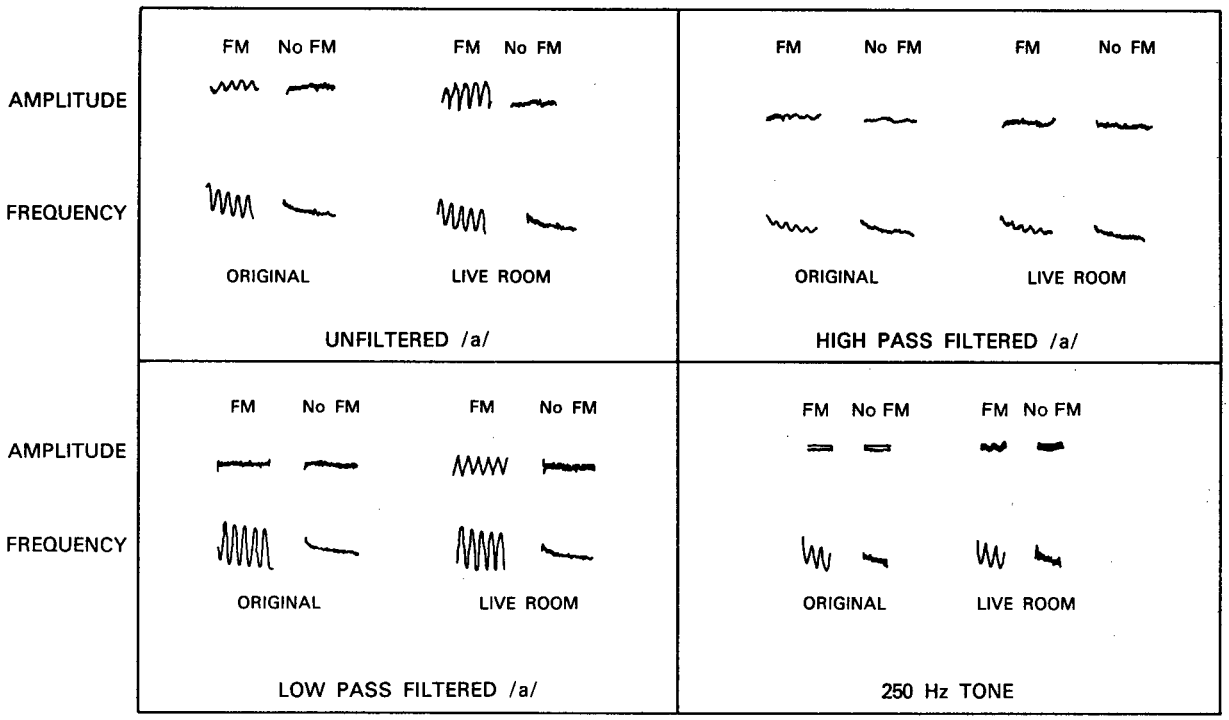

Fig. 1 — Amplitude and frequency tracings of original and live-room recordings for typical stimuli

large increase in amplitude fluctuation in the live room, and the low-passed vowels and the 250-Hz tone also had considerable amplitude fluctuations on the live-room tape. At least for the lower frequencies, the stimultaneous amplitude and frequency modulation probably accounts for the improved waver detection in the live-room environment. Performance on the live-room tapes was not as high as for the loudspeaker, but this could be caused by other factors, such as the lack of additional information from small head movements and diffractions around the head. It is possible that there are additional cues in the higher frequencies as well. The high-passed vowels showed improved detectability with the loudspeaker, in spite of the fact that the original small amplitude fluctuations disappeared in the live room recording.* The greater effect for the intact vowel may be caused by high-frequency cues in addition to the amplitude modulation.

A comparison of waver perception for the tones and for the vowels shows that small amounts of waver in f_0 were easily detected in the richer vowel stimuli, whereas greater pitch changes were required for the tones. The pitch of the vowel harmonics is modulated, as well as that of the fundamental, and in proportion to the number of the harmonic. This greater modulation of the harmonics may be used by the listeners to detect the vowel waver. The data for the high-passed and low-passed stimuli strongly suggest that this is the case. Even though different waver values were tested for the three sets of stimuli, it can be seen from Table 4 that waver detection for the high-passed vowels was very similar to that for the complete vowels, but a much larger Δf was required for detection for the low-passed vowels. Even so, the low-passed vowels required a smaller Δf than did the pure 250-Hz tone.

Table 6 shows the results for pairs where both vowels were wavered and indicates the discriminability of different amounts of waver. The results seem to indicate that discrimination was sometimes better with headphones and at other times with the loudspeaker. Two circumstances account for the situations where good discrimination occurred. Discrimination was high whenever the members of a pair straddled the detection threshold, yielding effectively one wavered and one unwavered stimulus. Good discrimination also occurred when both stimuli were above threshold and the difference in waver was relatively large. Waver discrimination is not necessarily better in the live-room environment, but it depends on where the stimuli lie relative to the waver detection threshold and on the relative difference in the extent of the waver.

CONCLUSIONS

Pitch waver in vowels and tones was more easily detected when the stimuli were heard over a loudspeaker in a live room or when a recording made in a live room was played over the headphones. The effect seems to be at least in part due to the occurrence of amplitude changes in addition to the frequency changes in the live-room environment. The ability to discriminate different degrees of waver was influenced by the waver threshold.

In spite of the perceptual differences between headphones and loudspeaker, preference data [11] indicated that subjects preferred Δfs of 1.8 to 2.2 Hz and f_ms of 2 to 4 Hz for both listening conditions. This means that the preferred Δf for loudspeaker listening would be well above threshold, but for headphones the preferred level is very near the detection threshold.

The effect was greater for intact vowels than for low- or high-passed vowels. The waver was more easily detected with the richer vowel stimuli than with pure tones. The relatively high modulation thresholds for the tones compared to those found by other investigators (e.g., Hartmann and Klein [12] or Zwicker [8]) were most likely the result of using naive, untrained listeners.

*The possibility that the original and the live-room measurements for the high-pass tapes might have been mislabeled was considered, but in view of the manner in which the measurements were made and of other internal consistencies, it is extremely unlikely that this is the case.

UNCLASSIFIED

Table 6 — Average A_g for Same-Different Judgments when Both Stimuli Were Modulated

Listening Condition	n	A_g for Contrast Pair Δf			
		0.70/1.1	0.70/1.8	1.8/2.8	1.1/2.8
Unfiltered /a/	15				
Loudspeaker		0.55	0.74	0.60	0.75
Live-room tape		0.68	0.83	0.53	0.71
Headphones		0.53	0.64	0.63	0.70
		0.44/1.1	0.44/2.8	2.8/7.0	1.1/7.0
Low-pass filtered	10				
Loudspeaker		0.62	0.87	0.67	0.83
Live-room tape		0.51	0.63	0.70	0.77
Headphones		0.50	0.63	0.68	0.80
		0.20/0.30	0.20/0.60	0.60/1.0	0.30/1.0
High-pass filtered	12				
Loudspeaker		0.59	0.51	0.50	0.70
Live-room tape		0.46	0.58	0.53	0.58
Headphones		0.56	0.50	0.54	0.62
		3.7/5.8	3.7/9.2	9.2/18.3	5.8/18.3
250-Hz tone	10				
Loudspeaker		0.52	0.80	0.67	0.84
Live-room tape		0.58	0.79	0.64	0.84
Headphones		0.50	0.74	0.71	0.88
		1.3/3	1.3/8	8/20	3/20
1000-Hz tone	11				
Loudspeaker		0.53	0.73	0.67	0.81
Live-room tape		0.56	0.83	0.60	0.82
Headphones		0.54	0.70	0.68	0.88

In a different area, it has been noted [13] that transient intermodulation distortion (TIM) that was easily heard over a loudspeaker could not be heard over headphones. The authors of Ref. 13 concluded that headphones were unsuitable for studying TIM and used only results for loudspeakers. Other perceptual phenomena may also demonstrate a loudspeaker-headphone difference.

Since much of normal listening occurs in live rooms, the investigation of perceptual phenomena should include the way in which the perceiver uses the additional cues that are available in real-world environments as well as the more rigorously controlled headphone environment.

ACKNOWLEDGMENTS

This work was supported by ONR under Task Area RR 021-05-42. We thank David C. Coulter for his assistance in setting up the equipment used to generate the wavered vowels and for making the amplitude and frequency measurements of the stimuli used in the third experiment.

We also thank Edith L. R. Corliss for her help in arranging for use of the anechoic chamber at the National Bureau of Standards, and special thanks are due to Charlie Mirachi for many hours of patient listening. We thank numerous colleagues and friends for their comments and ideas.

REFERENCES

1. J.L. Flanagan and M.G. Saslow, "Pitch Discrimination for Synthetic Vowels," *J. Acoust. Soc. Am.* **30**, 435-442 (1958).

2. H. Fastl, "Frequency Discrimination for Pulsed Versus Modulated Tones," *J. Acoust. Soc. Am.* **63**, 275-276 (1978).

3. E. Zwicker and G. Gässler, "Die Eignung des dynamischen Kopfhörers zur Untersuchung frequenzmodulierter Töne," *Acustica* **2**, AB134-AB139 (1952).

4. D.C. Coulter, personal communication, 1981.

5. G.E. Peterson and H.L. Barney, "Control Methods Used in a Study of the Vowels," *J. Acoust. Soc. Am.* **24**, 175-184 (1952).

6. I. Pollack, D.A. Norman, and E. Gallanter, "An Efficient Non-Parametric Analysis of Recognition Memory," *Psychonomic Science* **1**, 327-328 (1964).

7. D.M. Green, "General Prediction Relating Yes-No and Forced-Choice Results," *J. Acoust. Soc. Am.* **35**, 1042 (1964).

8. E. Zwicker, "Die Grenzen der Hörbarkeit der Amplitudenmodulation und der Frequenzmodulation eines Tones," *Acustica* **2**, AB125-AB133 (1952).

9. E. Zwicker and W. Kaiser, "Der Verlauf der Modulationsschwellen in der Hörfläche," *Acustica* **2**, AB239-AB246 (1952).

10. C.L. Ludlow, D. Coulter, and F. Gentges, "The Differential Sensitivity of Measures of Fundamental Frequency Perturbation to Laryngeal Neoplasms," in *Vocal Fold Physiology*, D. Bless and J. Abs, eds., University of Wisconsin Press, Madison, in press.

11. A. Schmidt-Nielsen and David C. Coulter, "Measurements of Waver Preference in Synthetic Vowels," NRL technical memorandum 7520-6, 1982.

12. W.M. Hartmann and M.A. Klein, "Theory of Frequency Modulation Detection for Low Modulation Frequencies," *J. Acoust. Soc. Am.* **67**, 935-946 (1980).

13. M. Petri-Larmi, M. Otala, E. Leinonen, and J. Lammasniemi, "Audibility of Transient Intermodulation Distortion," in *Proceedings of the 1978 IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1978, pp. 255-262.