EFFECTS OF ACOUSTIC CONTEXT ON DISCRIMINATION THRESHOLDS FOR FORMANT-LIKE TONE GLIDES.

Peter J. Bailey
Department of Psychology, University of York, York YO10 5DD, UK

ABSTRACT

Some perceptual characteristics of rising and falling third-formant-like tone glides were measured when the glides were presented in isolation, and in four more-or-less speechlike acoustic contexts. Glide levels at discrimination threshold were measured adaptively for five practised listeners in a two-interval procedure with feedback. The listeners' ability to categorise stimuli containing supra-threshold glides was also measured. If a phonetic mode of processing gave preferential access to information provided by the glides, lowest glide discrimination thresholds might have been expected for glides in the most speech-like context. However, thresholds were always lowest for glides in isolation. The data suggest that glide thresholds were more dependent on the masking effects of adjacent acoustic energy than on the likelihood that a stimulus was processed by a phonetic module.

1. INTRODUCTION

Synthetic tokens of the syllables /dɑ/ and /gɑ/ can be created in which the only distinctive acoustic property is the direction of the initial third-formant (F3) transition; in such syllables a falling F3 transition evokes a 'da' percept and a rising F3 transition evokes a 'ga' percept. Similar effects are obtained when the formant transitions are replaced by pure-tone glides. Such stimuli have been popular in 'duplex perception' experiments, in which the perceptual consequences of manipulating the level of the tone glides have been explored.

When the glide level is high relative to the formant levels in the syllable base (the syllable without a tone glide), listeners report a percept dominated by a whistle-like sound characteristic of a tone glide heard in isolation [9]; as the glide level is decreased a 'duplex' percept is reported of 'da' or 'ga' (depending on glide direction) together with the whistle. If the glide level is reduced further a 'duplexity threshold' is reached, when the percept becomes simply 'da' or 'ga' without the whistle [7]. If the glide is completely attenuated, leaving just the syllable base, the percept is typically of an ambiguous 'ga'. It has been reported that below the duplexity threshold there is a range of glide levels which can evoke 'da' and 'ga' percepts but at which the glides cannot be reliably detected or categorised as falling or rising [6,7]. This result has been taken as evidence that a phonetic mode of processing takes precedence over auditory processing of the tone glides.

However, the definition of duplexity threshold is critical to this interpretation of the data, and choosing an optimum procedure for measuring the glide level that, when mixed with the base syllable, corresponds to the duplexity threshold has not proved to be straightforward. Whalen and Liberman [7] used the method of adjustment and found that the mean glide levels at which a whistle was just audible were -6.4 dB and 0.0 dB for falling and rising glides respectively (relative to the level of F3 in the base). They showed further that although the syllable base with falling or rising glides at -10.4 dB could be identified reliably as 'da' or 'ga', glides at the same level in base context could not be matched reliably to rising or falling glides presented alone.

Bailey and Herrmann [1] used the method of constant stimuli in a two-interval two-alternative forced-choice procedure (2I2AFC) without feedback and found mean glide levels (corresponding to 75% correct performance) 13.1 dB and 10.5 dB lower (for falling and rising glides, respectively) than those reported by Whalen and Liberman. Bailey & Herrmann found that, relative to the duplexity threshold measured using their procedure, glides in base+glide syllables at levels corresponding to chance detection performance gave chance performance in identification tests.

Vorperian et al. [6] used an adaptive 2I2AFC procedure with feedback and found that the 71% correct mean thresholds were approximately 46 dB and 49 dB SPL for falling and rising glides presented in a base with overall level 83 dB SPL. Identification performance for base+glide syllables was maximal for glide levels between +10 dB and 0 dB relative to threshold, deteriorated systematically for glide levels below threshold, and reached chance performance for glides 20 dB below threshold. The decline in identification performance as glide levels decreased below threshold is broadly consistent with a decrease in glide detection accuracy from 71% correct at threshold to chance performance at some (unknown) sub-threshold level.

The approach taken by Xu et al. [9] was to avoid explicit threshold measurements altogether and use base+glide stimuli with falling or rising glides in which the glide level was varied systematically from near glide threshold to 60 dB above threshold. In one condition the listeners' task was to identify the stimuli as 'da' or 'ga', and in the other condition the task was to listen for the whistle and identify it as 'high' or 'low'. The results showed that maximal performance for the consonant identification task was found with glide levels 10 dB below the levels giving maximal performance for whistle identification.

It has not proved easy to provide an unequivocal demonstration that there is a glide level below which glides cannot be detected or discriminated but can nonetheless disambiguate the syllable base. This is in part because estimates of glide detection or discrimination thresholds vary according to the psychophysical procedures used for measuring them. Another contributing factor is the difficulty of establishing the role of differences in task and stimulus experience in the comparison of auditory and phonetic perceptual performance. Task demands have been shown to affect perceptual outcomes with speech stimuli [e.g.2], and listeners' long experience with hearing speech might enable them to respond to acoustically similar features differently in speech and in non-speech sounds. Considerations like these suggest it might be premature to take the results so far from experiments on duplex perception of base+glide syllables to indicate the perceptual priority of a phonetic module.

The experiment reported here was designed to contribute further data to this issue by measuring the effect on glide discrimination thresholds and perceptual categorisation when the acoustic context for tone glides was manipulated.

2. EXPERIMENT

2.1. Rationale and Predictions

It has been proposed that a phonetic module has privileged access to the auditory information that specifies phonetic identity, and that this is manifest as greater sensitivity for phonetic than for auditory processes [5,6,7]. It should therefore be possible to demonstrate that the discriminability of falling and rising tone glides varies with the extent to which they distinguish phonetic, as distinct from auditory, categories. In this experiment measurements of discrimination thresholds and categorisation performance were made for falling and rising F3-transition-like glides presented in isolation, and in four acoustic contexts. The conditions were: (a) glides + three-formant syllable base, (b) glides + syllable base excluding the F3 steady state, (c) glides + first and second formant (F1 and F2) transitions only, (d) glides + F3 steady state only, and (e) glides alone. The acoustic contexts of conditions (a) to (e) involve decreasing similarity to the acoustic patterns typically found in speech, and might be expected to be of decreasing interest to a phonetic module. If so, and assuming the phonetic module has perceptual priority, discrimination thresholds for falling and rising glides should increase across conditions (a) to (e), and, correspondingly, categorisation of the stimuli as ‘d-like’ or ‘g-like’ should be increasingly variable.

On purely auditory grounds, however, one should expect glide discrimination thresholds to depend primarily on the extent to which the glides are masked by the auditory context in which they appear. The main sources of masking for the tone glides are likely to be upward spread of masking from F1 and F2, and backward masking from F3 [3,4]. As backward masking is typically a weaker effect than upward spread of masking, auditory influences predict that glide discrimination thresholds should decrease across conditions (a) to (e).

2.2. Stimuli

A three-formant syllable base (referred to here as BF123) was synthesised using three second-order resonators in a parallel configuration with outputs summed in alternating phase. The excitation consisted of 40 equal-amplitude, cosine-phase harmonics of the fundamental, constrained so that for the F1 and F2 resonators the input was limited to harmonics 1–13, and for the F3 resonator to harmonics 14–40. The half-power bandwidths for the F1, F2 and F3 resonators were 50 Hz, 70 Hz and 100 Hz. Synthesis parameters for the base are shown in Table 1, where formant (Fn) and fundamental (F0) frequencies are given in Hz, overall amplitude (AO) in dB relative to maximum, and time in milliseconds. During the F1 and F2 transitions in the base F3 was fully attenuated and there was no energy above approximately 1625 Hz. The relative levels of F1 and F2 were equal and 10 dB above the level of F3 when it was present. Parameters were interpolated linearly between the values shown.

Three further stimuli were derived from the base BF123 to create the other acoustic contexts described above, as follows: (i) BF12: F3 fully attenuated throughout, (ii) BF12tr: all three formants fully attenuated after the initial transitions, and (iii) BF3: F3 steady state only, F1 and F2 fully attenuated throughout.

Nine F3 transitions were created with onset frequencies in the range 2750 Hz to 1750 Hz (in 125 Hz steps), terminal frequency 2375 Hz, relative level equal to that of F3 in the steady state of the base, and duration and amplitude envelope equal to those of the transitions in the base. A corresponding range of nine sinusoidal tone glides with similar spectro-temporal properties was also created by replacing the resonator in the synthesiser with a sinusoidal oscillator. The amplitude envelopes of all formant and tone glide onsets and offsets were shaped with 5 ms linear ramps.

ms	F0	AO	F1	F2	F3
0	125	-20	375	1600	
5	125	-10	375	1600	
65	125	0	700	1220	2375
245	110	-6	700	1220	2375
270	110	-20	700	1220	2375
275	110	-	700	1220	2375

Table 1. Synthesis parameters for the complete base BF123

2.3 Subjects

Five people took part in the experiment, which involved six listening sessions each of approximately one hour duration. Listeners were in the age range 20-30 years, and had normal pure-tone audiometric thresholds.

2.4 Procedure

All listening took place in a double-walled sound-attenuating room. Stimuli were synthesised using a sampling rate of 22.050 kHz, converted to voltages using 16-bit digital-to-analog converters, low-pass filtered at 6 kHz, and presented binaurally over headphones.

2.4.1 Categorisation. Six series of nine stimuli were presented to listeners for categorisation as ‘d’-like or ‘g’-like. Four series were created by mixing the nine tone glides with the base BF123 and its derivatives, BF12, BF12tr and BF3. The tone glides and contexts were mixed with synchronous onsets, and the glide level set 10dB below that of the F3 steady state in the base. The fifth series of stimuli consisted of the nine tone glides presented alone, at the same level as the glides in the other series. A sixth series was created by similarly mixing the nine F3 transitions with the base BF123, to create a series of complete 3-formant syllables. The stimulus levels were set such that the steady state portion of the base BF123 was at 70 dB SPL.

For each stimulus series, the nine stimuli were presented 10 times, in a new random order for each presentation. A trial began with a 100 ms visual cue presented 250 ms before the stimulus, and the next trial began 1500 ms after the listener’s response. The response, for which an unlimited amount of time was available, was a mouse click in one of two regions of the screeen labelled ‘D’ and ‘G’. The order of adminstration of the stimulus series was randomised for each listener, and listeners could take a break between each. All the categorisation data for each listener were collected in their first session.

2.4.2 Discrimination Thresholds. Glide levels corresponding to the threshold for discriminating the falling and rising tone glides with onset frequencies 2750 Hz and 1750 Hz were measured adaptively for glides presented alone and when mixed with each of the four base contexts, using a two-interval, two-alternative forced-choice procedure. A trial began with a 100 ms visual cue presented 250 ms before two stimulus intervals separated by 500 ms. One stimulus interval contained a falling tone glide and the other a rising tone glide, determined randomly. The interval containing the falling glide was arbitrarily designated the 'correct' interval, to which listeners responded with an appropriately-located mouse click. Feedback was given for 750 ms after every

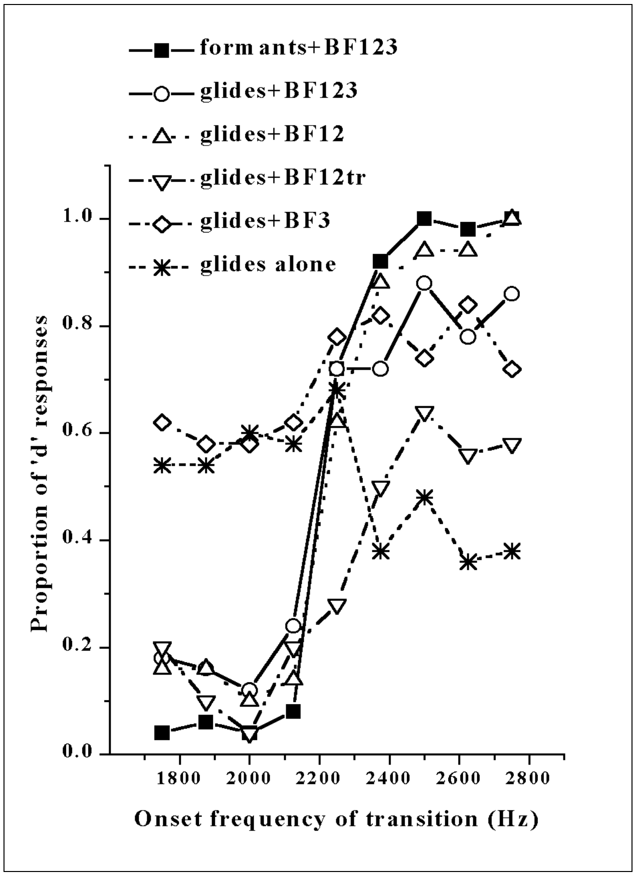

Figure 1. Categorisation functions averaged across listeners

response, for which time was unlimited, and the next trial followed 750 ms after the end of feedback. Additional feedback was continously available in the form of a number corresponding to the current glide level. For the conditions involving base contexts, the appropriate base was present in both intervals, and mixed with the glides so that the onsets of glide and base were synchronous. The level of the base contexts, when present, was controlled by a fixed attenuator set such that the steady state portion of the base BF123 was at 70 dB SPL.

Listeners were given no instructions about the task, beyond the requirement to maximise the number of correct responses, and to minimise the glide level in the course of a block of trials. Glide levels corresponding to 79% correct discrimination were determined using a three down, one up adaptive procedure. Sixteen threshold estimates were made, in two listening sessions, for each of the five conditions. Each listener started with a different condition, and completed 8 threshold estimates in each condition before cycling through the conditions again in the same order. The starting level of the glide was set to be approximately 15 dB above threshold, and the stepsize, initially 4 dB, was halved after the first and the third reversal. A threshold estimate involved a run of 60 trials, and thresholds were computed by omitting the first 3 or 4 reversals, to leave an even number, before averaging the levels at the remaining reversals. The median number of reversals contributing to a threshold was eight.

2.5 Results

2.5.1. Categorisation. The categorisation data are summarised in Figure 1, which shows the proportion of 'd-like' responses made to each of the stimuli in each stimulus series, averaged over the four listeners. The perceptual plausibility of the synthesis parameters for the base was evident in the consistent categorisation of complete syllables (formants+BF123) with falling formants as 'd'-like and those with rising formants as 'g'-like. Somewhat less consistent but nonetheless clear categorisation was possible for glides in the complete base (glides+BF123) and in the base without F3 (glides+BF12). Glide direction had a smaller effect on responses for glides in the transition-only base (glides+BF12tr) and in the F3-only base (glides+BF3); the former were categorised predominently as 'g'-like, and the latter as 'd'-like. Glides alone could not be categorised reliably with these labels, as demonstrated previously in [7] and in some conditions of [1].

2.5.2. Glide discrimination thresholds. Threshold was defined as the mean of those of the last eight threshold estimates that were within two standard deviations of the mean. Thresholds in dB SPL are shown in Figure 2 for each listener in each condition. The median number of estimates contributing to each threshold was

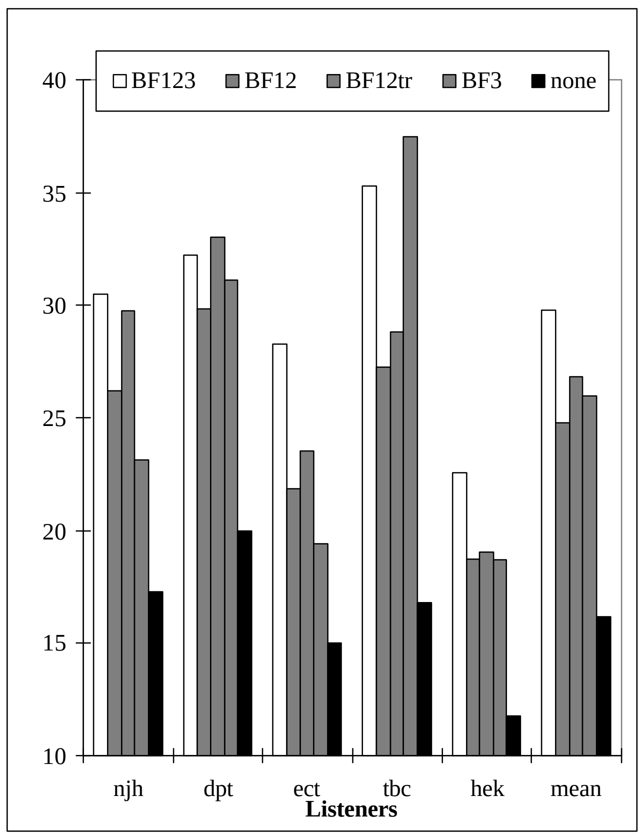

Figure 2. Glide thresholds in the contexts shown in the legend.

eight.

For all listeners the lowest discrimination threshold was obtained for glides presented alone. The highest mean threshold was obtained for glides in the complete base BF123. Thresholds for the glides in base context were higher but were not systematically different from each other. Pairwise comparisons among the means using the procedure recommended by Tukey showed that mean threshold for glides alone was statistically different from mean thresholds for all the other conditions ($p<0.01$), but that the other conditions did not differ from each other.

2.5.3. Relation between categorisation and threshold. The contribution of the reliability of categorisation to the glide discrimination threshold appeared to be small. For example, there was no reliable difference between the discrimination thresholds for glides in BF12 and BF3 contexts despite marked differences in categorisation for those conditions. The relationship between categorisation and threshold was quantified as follows. A sigmoidal function derived from the Boltzmann equation was fitted to the categorisation data for each subject in each condition. The width parameter from the fitted function, closely related to the slope of the function, was taken as an estimate of the consistency of categorisation. There was no relationship between these estimates of the consistency of categorisation and the glide discrimination thresholds, computed across all subjects and conditions ($r=0.094$, $N=25$).

2.6 Discussion

It was argued above that if a phonetic perceptual module takes precedence over auditory processing, as has been inferred from claims about the disambiguating power of apparently sub-threshold phonetically-relevant information, glide discrimination performance should be optimal for glides in a phonetic context. In the experiment reported here the phonetic context was supplied by a syllable base, which, when combined with falling or rising glides gave percepts reliably categorised as 'd'-like and 'g'-like. Falling or rising glides presented alone, and presumed therefore not to engage phonetic processing, were not categorised reliably as 'd' or 'g'.

A comparison of the estimates of glide discrimination thresholds for glides presented alone and in base BF123 context suggests that a phonetic context does not lead to better discrimination performance. On the contrary, performance was optimal, that is thresholds were lowest, when the glides were presented alone, and thresholds were highest for glides in the most speechlike context BF123[1]. Of course, notwithstanding the categorisation data described above, presenting glides in a speechlike base context does not guarantee that during threshold measurements glide+base stimuli will be processed by a phonetic module. However, the inclusion of detailed trial-by-trial feedback in the discrimination procedure should have ensured that listeners used the most sensitive perceptual strategy at their disposal.

The low thresholds reported here for isolated glides suggest that the primary determinant of discrimination performance was not the degree of engagement of a phonetic module, but rather the extent to which the auditory representations of the glides were masked by other acoustic energy present in the signal. The pattern of threshold estimates across conditions was not quite as auditory masking considerations might have predicted. Lowest thresholds were indeed found in the condition without additional masking energy, and highest thresholds in the condition with potentially the greatest masking energy. However there were no differences between the conditions in which masking effects might have been expected as a result of upward spread of masking from simultaneously present lower-frequency energy (as in conditions BF12 and BF12tr), and the conditions in which only the typically weaker effects of backward masking should have been evident (condition BF3). The similarity of thresholds across the different base conditions suggests that whatever cues listeners used to discriminate the glides near threshold must have been obscured by all versions of the base to a similar extent. Exploration of the origin(s) of the threshold elevation caused by the base contexts could prove to be generally informative about the representation of phonetic information in the complex auditory patterns of speech.

3. CONCLUSION

The results of this experiment suggest a need for caution in the interpretation of those data from experiments on duplex perception which have been taken to show that listeners are more sensitive to particular patterns of auditory information when they are phonetically relevant than when they are not.

ACKNOWLEDGMENTS

This work was supported in part by Grant No. G9306195N from the UK Medical Research Council. Nick Hill and Andrew Sach made useful suggestions, for which I am grateful.

NOTE

1. A different outcome was apparently obtained by Bentin and Repp (cited in [8]), who found similar discrimination thresholds for isolated glides and for glides in base context. In the absence of further details the difference in outcome of their experiment and this one is hard to rationalise.

REFERENCES

[1] Bailey, P. J. and Herrmann, P. 1993. A reexamination of duplex perception evoked by intensity differences. *Perception and Psychophysics,* 54, 20-32.

[2] Day, R. S. and Wood, C. C. 1971. Interactions between linguistics and non-linguistic processing. *Haskins Laboratories Status Report,* SR27, 185-192.

[3] Egan, J. P. and Hake, H. W. 1950. On the masking patterns of a simple auditory stimulus. *Journal of the Acoustical Society of America*, 22, 622-630.

[4] Elliot, L. 1971. Backward and forward masking. *Audiology,* 10, 65-76.

[5] Liberman, A. M. and Mattingly, I. G. 1989, A specialisation for speech perception. *Science,* 243, 489-494.

[6] Vorperian, H. K., Ochs, M. T. and Grantham, D. W. 1995. Stimulus intensity and fundamental frequency effects on duplex perception. *Journal of the Acoustical Society of America,* 98, 734-744.

[7] Whalen, D. H. and Liberman, A.M. 1987. Speech perception takes precedence over non-speech perception. *Science,* 237, 169-171.

[8] Whalen, D. H. and Liberman, A. M. 1996. Limits on phonetic integration in duplex perception. *Perception & Psychophysics,* 58, 857-870.

[9] Xu, Y. Liberman, A. M. and Whalen, D. W. 1997. On the immediacy of phonetic perception. *Psychological Science,* 8, 358-362.