Accepted after abstract only review

Mechanisms for the Perception of Concurrent Vowels

Hazel Moulder [1], [#] **Hamish Meffin** [2], **David Grayden** [2]

[1] Department of Otolaryngology, The University of Melbourne, Australia
[2] The Bionic Ear Institute, East Melbourne, Australia
[#] hmeffin@bionicear.org

Abstract

Previous research has shown that the perceptual segregation of concurrent vowels is improved when there is a difference in fundamental frequency (F0) between them. Two theories, termed F0-guided segregation and glottal pulse asynchrony (GPA), have been advanced to explain this "ΔF0 effect". A previous study found no consistent effect of GPA. However it is argued that this may have been because the auditory system uses both strategies, in which case a common F0 may cause the two vowels to be heard as one regardless of GPA. To overcome this potentially confounding influence, vowels with irregularly timed glottal pulses (and thus no well defined F0) were used to investigate the role of GPA. The results show that GPA still has no significant effect on recognition rates. Remarkably however, these irregularly excited vowels gave recognition rates that were equal to or significantly greater than their periodic counterparts, suggesting that F0-guided segregation is not required to explain the ΔF0 effect.

1. Introduction

Separating mixed speech signals is a significant feat, yet one which the auditory system performs comparatively easily. A large amount of previous research has indicated the benefit which a difference in fundamental frequency (F0) has on identification of simultaneous speech (Brokx and Nooteboom 1982). This is usually examined by presenting listeners with digitally-summed, synthesised vowel pairs and asking them to identify both components. Typically, a gradual improvement with increasing fundamental frequency separation (ΔF0) is observed up to around a semitone, after which performance asymptotes (Assmann and Summerfield 1990; Assmann and Summerfield 1994; Culling and Darwin 1993; Culling and Darwin 1994; de Cheveigné, Kawahara, Tsuzaki, and Aikawa 1997; Summerfield and Assmann 1991). Despite the clear replicability of this effect the mechanisms which are responsible for this gradual improvement are still unresolved.

The most obvious theory to explain the "ΔF0 effect", termed F0-guided segregation, holds that the pitch of each vowel is identified and used to distinguish the vowels when there is a difference in F0. This could occur in the spectral domain, for example by a harmonic sieve (Assmann and Summerfield 1990), or in the temporal domain by making use of the periodic nature of an individual vowel's waveform (de Cheveigné, Kawahara, Tsuzaki, and Aikawa 1997).

An alternative explanation, first noted by Summerfield and Assmann (Summerfield and Assmann 1991), relies on the observation that in the original experiments studying the ΔF0 effect, whenever the two vowels had the same F0, the glottal pulses were also (incidentally) synchronised (Fig.1A). They argued that this could lead to lower identification rates because, in contrast to the case in which the glottal pulses are asynchronous (Fig.1B), synchronous pulses would result in maximal interference between the waveforms of the two vowels. This would not be a problem when there is a difference in F0 between the vowels, because the glottal pulses of the two vowels will move in and out of alignment over time (Fig.1C), giving both asynchronous and synchronous pulses. It was then hypothesised that the auditory system is able to take advantage of those periods during which the pulses were asynchronous to best identify the vowels. Another factor that could make vowel pairs containing asynchronous glottal pulses easier to identify is that the auditory system may use the synchronous activity evoked across spectral channels by a single pulse as a cue to combine that activity into a single spectral estimate that pertains to just one vowel. As long as the pulses of the two vowels are asynchronous, they may be segregated by taking these "spectral snapshots" across frequency channels at the times of pulse-evoked waves of activity (Fig.1D).

Summerfield and Assmann (Summerfield and Assmann 1991) investigated whether subjects were able to use glottal pulse asynchrony (GPA, sometimes referred to as pitch period asynchrony) as a cue to segregate vowels, rather the differences in pitch via F0-guided segregation. This was achieved by using vowels with identical F0s but varying the degree of pulse asynchrony (0 or 1/2 of a pitch-period out of phase for F0=100Hz, and 0, 1/4, 1/2 or 3/4 of a pitch-period out of phase for F0=50Hz). No clear advantage was observed to be conferred by GPA in any condition except for dichotic presentation at 50Hz. This result appears to indicate that GPA is not generally used by the auditory system to segregate simultaneous vowels. However a problem with this interpretation arises if it is considered that the auditory system may use both GPA and F0-guided segregation to identify the vowels. In this case a vowel pair with the same F0 but asynchronous glottal pulses would provide conflicting cues regarding segregation. On the one hand, due to GPA there would be less interference between the two vowels and there would be evidence based on the "spectral snapshots" for the presence of two vowels. On the other hand, the presence of only a single F0 may be a very salient cue that only one vowel is present, since the likelihood of two vowels occurring with the exactly the same F0

Proceedings of the 10th Australian International Conference on Speech Science & Technology
Macquarie University, Sydney, December 8 to 10, 2004. Copyright, Australian Speech Science & Technology Association Inc.

Accepted after abstract only review

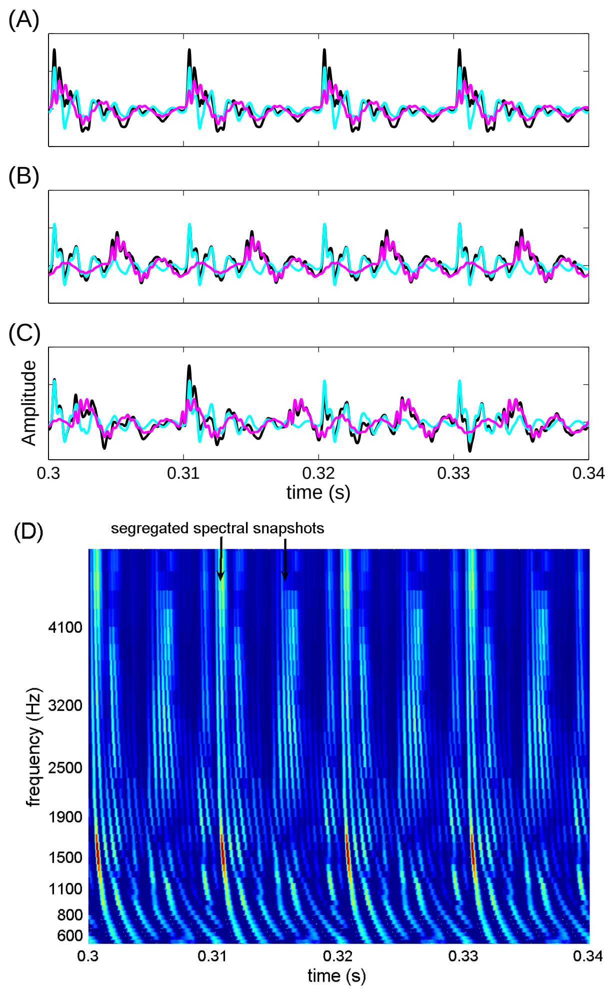

Figure 1: *(A)-(C) Waveforms of the synthetic vowels /æ/ (cyan) and /I/ (magenta) superimposed on top of the composite waveform (black). (A) Synchronous glottal pulses with F0=100 Hz for both vowels. (B) Asynchronous glottal pulses with F0=100 Hz for both vowels. (C) A difference in F0 of 26 Hz between the two vowels causes the glottal pulses of the vowels to move in and out of synchrony. (D) Spectrogram for the double vowel /æ/-/I/ derived from a model of auditory nerve response. The glottal pulses are asynchronous (as in (B)) allowing the possibility of making segregated "spectral snapshots" of each vowel.*

is very small. Under these circumstances it is possible that the strong salience of a single pitch cue wins out, and any segregation of the vowels based on GPA is lost because the auditory system recombines all the information in order to identify just a single vowel.

To test, in this study, whether the salience of a single pitch cue may have confounded Summerfield and Assmann's results by masking the role of GPA, we have devised vowel stimuli for which pitch salience is much reduced or absent, but which still allow GPA to be manipulated in systematic ways. This was done by creating stimuli in which glottal pulses were not periodic but whose inter-pulse interval (IPI) varied randomly about some mean. Three degrees of IPI irregularity were considered: periodic, in which the IPI is constant (=10 ms) and corresponds to the normal situation with F0=100 Hz; jittered, in which IPI varied randomly between 9.0 ms and 11.0 ms; and random, in which IPI varied randomly between 7.5 ms and 12.5 ms. Vowels in a pair were either both periodic, both jittered or both random. It was hypothesised that (Hypothesis 1:GPA) introducing glotal pulse asynchrony would improve identification of vowel pairs with jittered and random IPIs but not with periodic IPIs. To confirm that the pitch salience of the jittered and random IPI stimuli was reduced or absent and that the auditory system cannot not use mean IPI as a cue for segregation in random IPI stimuli, control experiments were run in which the vowels had a difference in mean IPI (ΔmIPI). It was hypothesised that (Hypothesis 2:ΔmIPI) introducing a ΔmIPI, will improve identification rates of vowels pairs with periodic IPIs, but not with random IPIs, when compared to pairs with GPA and the same mIPI.

2. Methods

2.1. Stimuli

Vowels were synthesised using a Klatt synthesiser (Klatt 2000) (44100 Hz sample rate, 32 bit amplitude quantization) implemented in Matlab Simulink by Sean McLennan (Lennan and Kewley-Port 2000). Two sets of five vowels were used: one as a training set to allow subjects to gain familiarity with the task, and a second, distinct test set that was used to obtain the data presented here. Table 1 lists formant frequencies and bandwidths used for the vowels in each set.

To test the first hypothesis, on GPA, vowels with the same mIPI of 10 ms were used, but with three degrees of IPI irregularity: periodic, in which the IPI =10 ms; jittered, in which IPI varied randomly between 9.0 ms and 11.0 ms; and random, in which IPI varied randomly between 7.5 ms and 12.5 ms. Four conditions of GPA were used corresponding to 0, 1/4, 1/2 or 3/4 of an IPI. The details of the procedure for generating vowel pairs is described in the Appendix.

To test the second (control) hypothesis, on mIPI, vowel pairs were also synthesized with a ΔmIPI equivalent in the periodic case to $1/4, 1/2, 1, 2$ or 4 semitones difference. This gave one vowel with a mIPI of 10ms and the second with a mIPI of 9.86, 9.72, 9.44, 8.91 and 7.94 ms, respectively. The procedure involved synthesizing two vowel pairs according to the algorithm described in the Appendix, one pair with a mIPI of 10 ms and the other pair with the lower mIPI. Then vowel 1 from the first pair and vowel 2 from the second pair were combined to obtain the double vowel, which ensured that both vowels had the same form of IPI distribution. The particular choice of GPA was not important in this process as it changed throughout the stimuli anyway, so 0 GPA was chosen arbitrarily.

Paired vowels were either both periodic, both jittered or both random. Thus the total double vowel stimulus set consisted of (10 vowel pairs) × (3 IPI irregularity conditions)× ((4 GPA conditions) + (5 mIPI conditions)), giving a total 270 distinct stimuli for both the training and test sets. The rms amplitude of vowels was normalized before they were summed to form a pair. Each pair was 1000 ms in duration plus another 5 ms linear ramp at the beginning and end of the stimuli. This duration was chosen to ensure that the

Accepted after abstract only review

Table 1: *Synthesis parameters for the vowels. F4=3300, B4=200, F5=3750, B5=250, F6=4900 and B6=1000 Hz for all vowels.*

Vowel	Label (example)	F1 (Hz)	B1 (Hz)	F2 (Hz)	B2 (Hz)	F3 (Hz)	B3 (Hz)
Training Set							
/i/	ee (beet)	254	50	2295	200	2989	400
/3/	ir (bird)	517	100	1329	60	1670	110
/u/	oo (boot)	254	65	889	110	2198	140
/U/	oo (foot)	430	80	1065	100	2198	80
/⊃/	aw (bawd)	517	90	889	100	2373	80
Test Set							
/æ/	a (bad)	693	70	1768	150	2373	320
/ε/	e (bed)	517	60	1856	90	2461	200
/I/	i (bid)	342	50	2032	100	2549	140
/a/	ar (bard)	693	130	1065	70	2461	160
/∧/	u (bud)	605	80	1241	50	2373	140

shifting pattern of GPA had sufficient time to pass through at least one beat.

2.2. Procedure

Testing was performed in a sound-attenuating booth. Stimuli were played using headphones (Phillips Electret N6325) presented by the computer. Presentation level was calibrated at 65dB A(peak) for the vowel /∧/. Subjects attended both a training session and a testing session, each approximately one and a half hour long. Subjects could have a break at any stage of the experiment as desired.

Training Session: Hearing was tested in the ear of presentation. Given a normal result, all ten vowels were presented in the periodic, jittered and random conditions. After this, participants were free to press any of the five test vowel buttons to hear one of its variations, until they felt competent at identifying them. Then a program presented these 30 stimuli randomly, requiring a button to be pressed according to which vowel was perceived. A score over 90 percent correct (with up to four attempts) allowed the subject to proceed to double vowel training, in which the 270 randomly ordered training vowel pairs were presented via a computer program in two blocks of 135 with immediate corrective feedback after each vowel pair.

Testing Session: The testing session was similar to the training session, except in the following details. Hearing was not retested. Only the five vowels from the test set were used and tested in the single vowel phase, with each stimulus presented 7 times giving 105 presentations, with a score above 90 percent required. Double vowel test stimuli were presented without feedback, and in 15 blocks of 54, to give 810 randomly-presented stimuli. Participants could hear single vowel stimuli again between blocks.

2.3. Subjects

The participants were nine, uncompensated, adult volunteers, recruited by open-invitation e-mail to staff and students in the precinct, or by personal invitation from the experimenters. None had prior experience with a double-vowel task. All were competent speakers of English, although for two, English was their non-native tongue. Six were female and three male.

3. Results

The mean scores for both vowels correct are shown in Fig.2 for the periodic, jittered and random conditions as a function of GPA and ΔmIPI, averaged across vowels pairs and subjects. The four GPA conditions (0, 1/4, 1/2 and 3/4 of an IPI) are shown in the left hand side of the plot against a grey background, while the five non-zero ΔmIPI conditions (1/4, 1/2, 1, 2 and 4 equivalent semitones) are shown in the right hand side of the plot against a white background.

3.1. GPA hypothesis

Focusing on the four GPA conditions in the left hand side of Fig.2, there does not appear to be any significant effect of GPA on vowel recognition rates for either periodic, jittered or random vowel-types. This contradicts our hypothesis that GPA would improve the recognition of double vowels for the jittered and random conditions, but not for the periodic condition. This observation was confirmed by performing a 2-way ANOVA on the scores for both vowels correct, averaged across the 10 vowel pairs, with factors GPA and SUBJECT and with the data split across the periodic, jittered and random conditions. For each of the three conditions of IPI irregularity, only SUBJECT emerged as a significant factor ($p < 0.0005$, see Fig.4) while GPA was not significant ($p = 0.76, 0.22, 0.36$ for the periodic, jittered and random conditions, respectively). This is consistent with GPA having no effect on double vowel recognition.

While a small effect of GPA cannot be excluded on the basis of these data, they are inconsistent with the hypothesis that GPA can explain the classic ΔF0 effect. Such a hypothesis would require (1) a common baseline for recognition rates across the periodic, jittered and random conditions when there is no GPA (GPA = 0 IPI) and (2) that the improvement in recognition rates when GPA is introduced for the jittered and random conditions is equivalent to the improvement seen in the classical ΔF0 effect (i.e. for the periodic condition). The data fail to satisfy the first part of this requirement. Under a 2-way ANOVA with only the data from GPA =0 IPI, and with IPI IRREGULARITY (i.e. periodic, jittered, random) and SUBJECT as factors, IPI IRREGULARITY is a significant factor ($p < 0.001$). *Post*

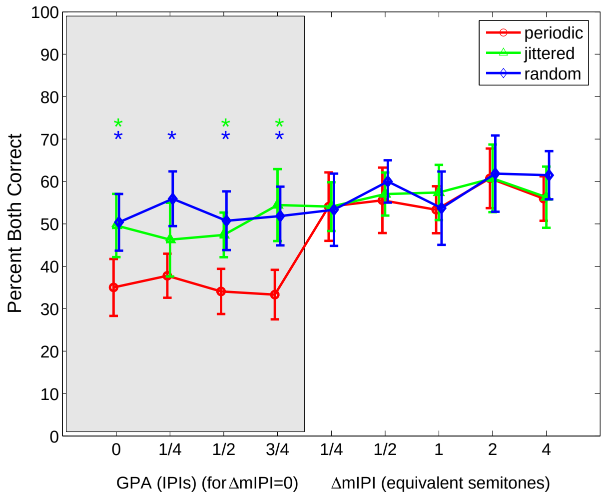

Figure 2: *Mean scores for both vowels correct for the periodic, jittered and random conditions as a function of GPA and* Δ*mIPI, averaged across vowels pairs and subjects. The four GPA conditions are shown in the left hand side of the plot against a grey background, while the five non-zero* Δ*mIPI conditions are shown in the right hand side of the plot against a white background. Error bars show* ±*one standard error. Colour coded asterisks indicate those conditions for which the scores for jittered (green) or random (blue) double vowels were significantly greater than the corresponding periodic case (at an alpha level of 0.05). Otherwise such pairwise differences were not significant.*

hoc Tukey pairwise comparisons revealed that the recognition rates for the periodic condition are significantly different from the jittered and random conditions (at an alpha level of 0.05), thus violating this requirement.

3.2. ΔmIPI hypothesis

To test the second hypothesis, that ΔmIPI would improve identification of double vowels for periodic but not jittered or random stimuli, a similar 2-way ANOVA was performed, but with factors ΔmIPI and SUBJECT. The factor ΔmIPI consisted of eight conditions: five conditions with ΔmIPI > 0 (1/4, 1/2, 1, 2 and 4 equivalent semitones) together with the three conditions ΔmIPI =0 and GPA=1/4, 1/2 and 3/4. The data were again split across the periodic, jittered and random conditions. The analysis shows that ΔmIPI is a significant factor for all three degrees of IPI irregularity (with $p < 0.001, 0.002, 0.002$, for the periodic, jittered and random conditions, respectively).

Tukey pairwise comparisons were performed. For the periodic condition, two significantly distinct subgroups were identified: one group consisted of those conditions with ΔmIPI =0 and a second one comprised those with ΔmIPI> 0. This result is a classic ΔF0-type effect, but with a sudden transition from ΔmIPI =0 to ΔmIPI> 0, instead of the usual gradual improvement up to 1 semitone.

For the jittered and random conditions, distinct subgroups do not emerge. The tests show significant differences between conditions with GPA = 1/4 or 1/2 IPIs (and ΔmIPI =0) and the condition with ΔmIPI= 2 equivalent semitones, but no other pairwise differences were signifi-

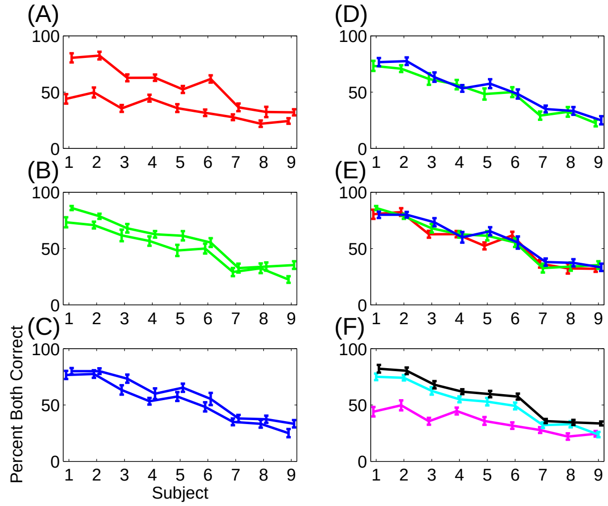

Figure 3: *Comparison of data across subjects leading to the simplified grouping of conditions as described in the text (section 3.4.). Red, green and blue lines signify the periodic, jittered and random conditions. (A)-(C) Comparison between the mean scores averaged over* Δ*mIPI = 0 conditions (lower trace) and* Δ*mIPI*> 0 *conditions (upper trace), for periodic (A), jittered (B) and random (C) cases. In each plot a significant difference exists that is consistent across subjects. (D) Comparison of the jittered and random scores averaged across* Δ*mIPI = 0 conditions, showing no significant differences. (E) Comparison of the periodic, jittered and random scores averaged across* Δ*mIPI*> 0 *conditions, showing no significant differences. (F) Comparison of the three grouped conditions described in section 3.4.. From lower to upper trace the grouped conditions are* $c = 1$ *(magenta),* $c = 2$ *(cyan) and* $c = 3$ *(black). Significant difference exist between all three conditions, which are consistent across subjects. Error bars show* ±*one standard error.*

cant. In summary, the results of the ANOVA for the jittered and random conditions indicate that the value of ΔmIPI does affect double vowel recognition, in contradiction to our second hypothesis, but the results of the pairwise comparisons indicate that this is not a very strong or systematic effect.

3.3. F0-guided segregation

A remarkable aspect of the data presented in this study is that double-vowel recognition rates for jittered and random vowels are greater than or equal to their periodic counterparts across all conditions. This is shown in Fig.2 where green (respectively blue) asterisks indicate conditions for which scores for the jittered (random) vowels were significantly greater than for periodic vowels in a Tukey pairwise comparison test ($p < 0.05$). For all other conditions there was no significant difference between the jittered/random and periodic cases. These observations are inconsistent with the hypothesis that an F0-guided strategy can explain the classic ΔF0-effect, since such a strategy would be expected to give significantly worse rates of recognition for jittered and random stimuli.

Accepted after abstract only review

3.4. A simple statistical model

A simple statistical model, requiring just two variables, captures the main significant effects observed in the data, including subject variability. In all it accounts for nearly 70 percent of the variance in the data. The model supposes that a subject's mean score, $\mu_{s,c}$, in any given condition is given by the product of two factors,

$$\mu_{s,c} = g_s f_c, \quad (1)$$

where g_s is a factor depending on the subject, s, and f_c is a factor depending on the condition, c. Given this premise, the least mean squares fit to data occurs when the factor g_s is the mean score of a subject (over all conditions) and the factor f_c is the mean score for a given condition (over all subjects) as a ratio of the global mean (over all subjects and conditions).

The number of conditions for f_c was reduced from a total of 27 (i.e., [3 IPI irregularity conditions] $\times$ ([4 GPA conditions] + [5 ΔmIPI conditions])) to just three grouped conditions required to describe the main effects. This is based on Fig. 2 which shows f_c (up to the constant global mean factor) for the 27 conditions, since it plots the mean score over all subjects by condition. From this the following three grouped conditions were identified, for which the mean score within the group was assumed to be the same for the model: (1) Conditions with ΔmIPI = 0 and periodic IPIs, regardless of GPA (the red data points in the grey area in Fig.2). (2) Conditions with ΔmIPI = 0 and jittered or random IPIs, regardless of GPA (the green and blue data points in the grey area in Fig.2). (3) All conditions with ΔmIPI > 0, regardless of whether the vowels were periodic, jittered or random (the red, green and blue data points in the white area in Fig.2).

It might be suggested that this scheme could be simplified one step further by combining the last two grouped conditions (2 & 3) into one. However there are significant differences between these two grouped conditions that are apparent even on a subject-by-subject basis as shown in Fig. 3A-C. In the first column, the mean scores over conditions with ΔmIPI = 0 and ΔmIPI > 0 have been calculated separately and plotted for each subject for periodic, jittered and random cases. For the jittered (Fig. 3B) and random (Fig. 3C) stimuli there is a small, but significant difference between the two conditions which is consistent across subjects (2-way ANOVA with $p < 0.001$) (unsurprisingly there is also a significant and consistent difference in the periodic case (Fig. 3A)). It is also important to justify combining the jittered and random stimuli in grouped condition 2, and combining the periodic, jittered and random stimuli in grouped condition 3. The rationale for this can be seen in Fig.3D and E respectively where five of the lines from Fig. 3A-C are replotted to emphasize these similarities. In Fig.3D the mean scores over conditions with ΔmIPI = 0 are shown for jittered and random double vowels by subject. A 2-way ANOVA with JITTERED/RANDOM and SUBJECT as factors reveals that the jittered and random conditions are not significantly different at an alpha level of 0.05. Similarly Fig.3E shows mean scores over conditions with ΔmIPI> 0 for the periodic, jittered and random conditions by subject and 2-way ANOVA reveals no significant difference between these three degrees of IPI irregularity. Finally, in Fig.3F, the scores for grouped conditions 1, 2 and 3 are shown in black, cyan and magenta (respectively) by subject. A 2-way ANOVA with GROUPED CONDITION and SUBJECT as factors shows that GROUPED condition is highly significant ($p < 0.001$) and *post hoc* Tukey pairwise comparisons show that each of the grouped conditions, 1, 2 and 3, are significantly different from each other at an alpha level of 0.05.

As a further test, the model predictions vs. observed data were subjected to a regression analysis. A high degree of correlation was found (Pearson correlation = 0.83) which was highly significant ($p < 0.001$). An R^2 value of 0.69 was obtained indicating that the model accounted for 69 percent of the variance in the data.

4. Conclusion

In this study we have considered whether a previously noted absence of an effect of GPA on double vowel segregation (Summerfield and Assmann 1991) may have been due to a masking effect brought about because the two vowels had the same F0. In the introduction we argued that such a salient pitch cue may cause the auditory system to treat the two vowels as one, thus suppressing the role of GPA. To test this hypothesis, vowels with reduced or eliminated pitch salience were synthesised by making the timing of glottal pulses irregular instead of periodic. These vowels were then used to investigate the role of GPA without the potentially confounding effects of a common F0. The conclusions of the study are encapsulated quantitatively in the simplified model in section 3.4. which shows that they are largely independent of subject. They may be summarised as follows.

- **GPA hypothesis** No significant effect of GPA was observed in the data for periodic, jittered or random stimuli, in contradiction to the stated hypothesis. The data are inconsistent with GPA playing a large role in the classical ΔF0 effect.

- **ΔmIPI hypothesis** Significant effects of ΔmIPI were observed for periodic, jittered or random stimuli. For the periodic stimuli, this was manifest as a ΔF0-like effect, but with a sharp improvement in score from ΔmIPI=0 to ΔmIPI> 0 instead of the normally observed gradual improvement. For the jittered and random stimuli, the effect was much smaller and more difficult to identify with the present data, but is consistent with a small, sharp improvement in score from ΔmIPI=0 to ΔmIPI> 0.

- **F0-guided segregation** Recognition rates for jittered or random stimuli were either equal to or significantly greater than those for periodic stimuli. This is consistent with F0-guided segregation not playing a large role in the classical ΔF0 effect, since otherwise significantly worse recognition rates for jittered and random stimuli would be predicted.

5. Appendix

Double vowels with the desired GPA were obtained by creating pulse trains P_1 and P_2 for each vowel in a pair in

Accepted after abstract only review

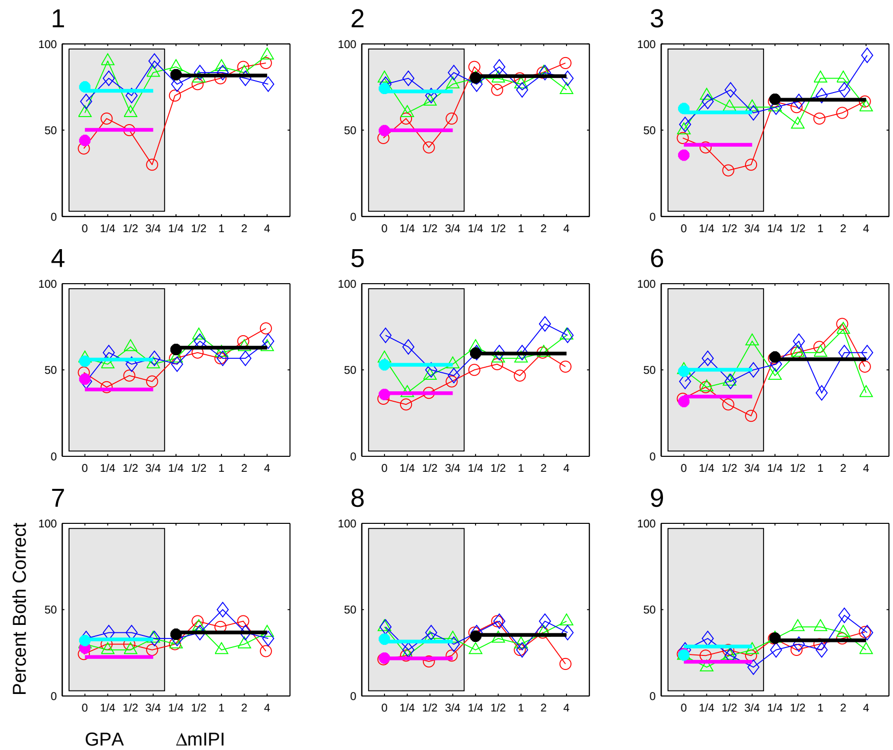

Figure 4: *Plots analogous to Fig.2 for each of the nine subjects, but also showing comparison to the predictions of the model described in section 3.4.. The scores for each subject as a function of GPA and* Δ*mIPI are shown for the periodic (red), jittered (green) and random (blue) cases. The predictions for grouped conditions* $c = 1$ *(magenta),* $c = 2$ *(cyan) and* $c = 3$ *(black) are shown as solid lines overlaid on the appropriate portion of the plot. Also shown is the true mean for each grouped condition for that subject as a filled circle of the appropriate color.*

the following manner. The first pulse of P_1 was at $t = 0$ ms. Thereafter pulses were added to P_1 and P_2 at times t_1 and t_2 in an alternating fashion beginning with P_2, and according to the updating equations

$$t_2 = t_1 + d_2, \quad (2)$$
$$t_1 = t_2 + d_1. \quad (3)$$

d_1 (d_2) were random time increments chosen from uniform distributions on the interval $[\mu_1 - \sigma, \mu_1 + \sigma]$ ($[\mu_2 - \sigma, \mu_2 + \sigma]$), where the mIPI was $\mu_1 + \mu_2 = 10$ ms, and $\sigma = 0, 0.5, 1.25$ ms for periodic, jittered and random conditions, respectively. The four conditions of GPA, viz. 0, 1/4, 1/2 and 3/4 of an IPI, where obtained by choosing $\mu_2 = 0.0, 2.5, 5.0$ and 7.5 ms, respectively (corresponding to $\mu_1 = 10.0, 7.5, 5.0$ and 2.5 ms). For vowels with the shorter mIPI, σ was scaled proportionately.

6. Acknowledgements

The authors thank two anonymous referees for their comments.

References

Assmann, P. F. and Q. Summerfield (1990). Modelling the perception of concurrent vowels: Vowels with different fundamental frequencies. *J. Acoust. Soc. Am. 88*, 680–697.

Assmann, P. F. and Q. Summerfield (1994). The contribution of waveform interactions to the perception of concurrent vowels. *J. Acoust. Soc. Am. 95*, 471–484.

Brokx, J. P. L. and S. Nooteboom (1982). Intonation and the perceptual separation of simultaneous voices. *J. Phonetics. 10*, 23–36.

Culling, J. F. and C. J. Darwin (1993). Perceptual separation of simultaneous vowels: within and across-formant grouping by f0. *J. Acoust. Soc. Am. 93*, 3454–3467.

Culling, J. F. and C. J. Darwin (1994). Perceptual and computational separation of simultaneous vowels: cues arising from low-frequency beating. *J. Acoust. Soc. Am. 95*, 1559–1569.

de Cheveigné, A., H. Kawahara, M. Tsuzaki, and K. Aikawa (1997). Concurrent vowel identification, I. effects of relative amplitude and F0 difference. *J. Acoust. Soc. Am. 101*, 2839–2847.

Klatt, D. (2000). Software for a cascade / parallel formant synthesizer. *J. Acoust. Soc. Am. 67*, 971–995.

Lennan, S. M. and D. Kewley-Port (2000). Klatt synthesizer in simulink.

Summerfield, Q. and P. F. Assmann (1991). Perception of concurrent vowels: Effects of harmonic misalignment and pitch-period asynchrony. *J. Acoust. Soc. Am. 98*, 1364–1377.