

# Temporal Integration of Consecutive Tones Into Synthetic Vowels Demonstrates Perceptual Assembly in Audition

Jefta D. Saija  
University of Groningen,  
University Medical Center Groningen

Tjeerd C. Andringa  
University of Groningen

Deniz Başkent  
University of Groningen,  
University Medical Center Groningen

Elkan G. Akyürek  
University of Groningen

Temporal integration is the perceptual process combining sensory stimulation over time into longer percepts that can span over 10 times the duration of a minimally detectable stimulus. Particularly in the auditory domain, such “long-term” temporal integration has been characterized as a relatively simple function that acts chiefly to bridge brief input gaps, and which places integrated stimuli on temporal coordinates while preserving their temporal order information. These properties are not observed in visual temporal integration, suggesting they might be modality specific. The present study challenges that view. Participants were presented with rapid series of successive tone stimuli, in which two separate, deviant target tones were to be identified. Critically, the target tone pair would be perceived as a single synthetic vowel if they were interpreted to be simultaneous. During the task, despite that the targets were always sequential and never actually overlapping, listeners frequently reported hearing just one sound, the synthetic vowel, rather than two successive tones. The results demonstrate that auditory temporal integration, like its visual counterpart, truly assembles a percept from sensory inputs across time, and does not just summate time-ordered (identical) inputs or fill gaps therein. This finding supports the idea that temporal integration is a universal function of the human perceptual system.

*Keywords:* temporal integration, synthetic vowels, rapid serial auditory presentation

Stimulus detection thresholds and stimulus duration are inversely related. In other words, the threshold for detecting an auditory stimulus decreases when its duration increases. For normal hearing listeners, each 10-fold in duration corresponds on average to a threshold drop of 8 to 10 dB (Hughes, 1946; Plomp & Bouman, 1959), and this relation holds for stimulus durations of a few hundred ms. When stimulus intensity is held constant (Munson, 1947), the perceived loudness of a tone increases gradually from onset until a steady loudness is reached at a certain duration. These effects are often described as the temporal integration of acoustic energy. It is usually

modeled as a leaky integrator (cf. Viemeister & Wakefield, 1991) that sums up acoustic energy over time within frequency bands, but leaks energy exponentially (Plomp & Bouman, 1959; Zwillocki, 1960). Various models of temporal integration have been proposed in terms of electric circuits (Jeffress, 1967; Munson, 1947) and neural excitation (Zwillocki, 1960). These models usually assume a relatively long-temporal window of about 200 ms, a duration in line with psychophysical observations, which make these models perfect for explaining integration phenomena such as threshold reduction and loudness augmentation.

This article was published Online First December 23, 2013.

Jefta D. Saija, Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, University of Groningen, The Netherlands, and Faculty of Mathematics and Natural Sciences, Artificial Intelligence and Cognitive Engineering, Department of Psychology, Experimental Psychology, and Research School of Behavioral and Cognitive Neurosciences, University of Groningen, The Netherlands; Tjeerd C. Andringa, Faculty of Mathematics and Natural Sciences, Artificial Intelligence and Cognitive Engineering and Research School of Behavioral and Cognitive Neurosciences, University of Groningen, The Netherlands; Deniz Başkent, Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, University of Groningen, The Netherlands, and Research

School of Behavioral and Cognitive Neurosciences, University of Groningen, The Netherlands; Elkan G. Akyürek, Department of Psychology, Experimental Psychology and Research School of Behavioral and Cognitive Neurosciences, University of Groningen, The Netherlands.

Part of this research was supported by a Rosalind Franklin Fellowship from the University of Groningen, University Medical Center Groningen and a VIDI grant (Grant No. 016.096.397) from the Netherlands Organisation for Scientific Research (NWO) and the Netherlands Organization for Health Research and Development (ZonMw) awarded to Deniz Başkent.

Correspondence concerning this article should be addressed to Jefta D. Saija, Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, University of Groningen, PO Box 30.001, 9700 RB Groningen, The Netherlands. E-mail: j.d.saija@umcg.nl

## Multiple Stimuli in One Memory Trace

Recent studies on auditory sensory memory have supported the idea that auditory stimuli are integrated over such comparatively long-time intervals. In this field, many studies have used electroencephalography to measure a component of the event-related potential called the mismatch-negativity (MMN). The presence of an MMN after stimulus presentation means that a violation of the norm in a series of stimuli is perceived. Any deviation in a to-be-expected order or identity of sequential stimuli can elicit an MMN, including deviations from preceding stimuli that are represented by a short-term memory (STM) trace in the auditory cortex (for reviews, see Näätänen, Kujala, & Winkler, 2011; Näätänen, Paavilainen, Rinne, & Alho, 2007).

An MMN study by Tervaniemi, Saarinen, Paavilainen, Danilova, and Näätänen (1994), who investigated the effect of deviations in tone pairs, suggested that two closely spaced stimuli, with an interstimulus interval (ISI) of maximally 140 ms, can be integrated into a single unitary sensory event. Yabe et al. (1998) came to a similar conclusion while investigating the effect of stimulus omission in trains of stimuli with different stimulus onset asynchronies (SOAs) on MMN responses obtained with magnetoencephalography, and estimated the temporal integration window to be around 160 to 170 ms (cf. Yabe et al., 1998). Others have estimated this to be slightly longer, around 200 ms (Sussman, Winkler, Ritter, Alho, & Näätänen, 1999).

In their influential review paper, Näätänen and Winkler (1999) concluded that auditory temporal integration is not merely a process of reducing auditory noise by compressing the time dimension (Näätänen, 1995), such as bridging a small gap or summing up energies, but is rather a constructive process that combines auditory information (pitch, loudness, duration, location, and energy) into a single perceptual event. This idea is also consistent with the larger concept of auditory scene analysis, a general model of auditory perception in which signal components that are produced by the same source are perceptually grouped into the same auditory objects (Bregman, 1994).

More important, Näätänen and Winkler (1999) proposed that an auditory episodic memory trace is established when combined input from different acoustic feature detectors is placed on “temporal coordinates” (i.e., preserving temporal order information within the trace). The authors posited a parallel between the medium of space, which is central to visual feature integration (e.g., Treisman, 1996), and that of time, which is central to auditory integration. Only after this temporal trajectory is established does the memory trace constitute a genuine acoustic object that can be perceived and experienced subjectively. The formation of these object representations is assumed to occur within a continuous sliding temporal integration window of about 200 ms (Näätänen, 1990 as in Näätänen & Winkler, 1999), although the temporal window of integration might also start at stimulus onset (Yu et al., 2011). Either way, this conceptualization of temporal integration in audition seems like a free lunch: Forming an integrated percept while fully preserving all temporal information suggests that temporal integration is costless in terms of maintaining the properties of the input signal. The current study sought to investigate this claim because there is evidence to the contrary from visual paradigms.

## Similarities to Vision

Assuming that auditory and visual perception operate on similar principles, studies on visual temporal integration may provide important insights into auditory temporal integration. In the so-called missing element task (MET; Akyürek, Schubö, & Hommel, 2010), observers view stimuli that are arranged in an evenly spaced square grid, across two successive partial displays (e.g., Hogben & Di Lollo, 1974). For instance, using a grid of 25 positions ( $5 \times 5$ ), observers are first shown a set of 12 stimuli, and then another set of 12 (i.e., 24 in total). Observers are asked to locate the one remaining empty position. Finding the missing element is virtually impossible by mentally comparing and examining the two stimulus displays. When those two displays are temporally integrated, however, they appear as if they were overlaid and then the missing element is immediately apparent. Because temporal integration is more likely to occur at shorter SOAs, the typical finding in the MET is that shorter SOAs result in higher task performance. Evidence from the MET shows that although information about individual parts appears to be inaccessible, the sum still is and constitutes the integrated percept. This contrasts with the findings from the previously discussed auditory studies, which suggested that information about individual parts can be accessed while also being combined into an integrated percept.

Further data on the nature of visual temporal integration has been obtained in studies that investigated performance in dual-target rapid serial visual presentation (RSVP) tasks. In such tasks two targets (T1 and T2) of short duration are presented among distractors in rapid succession (often with short blank gaps in between stimuli), and the participant is asked to report the identity and order of the targets. T2 can follow T1 with or without distractors in between and this distance is denoted as lag. Lag 3 for example means that T2 follows T1 with two distractors in between, thus T2 lags T1 as the third item. In RSVP tasks, participants often fail to report T2 when it follows T1 closely, within  $\pm 500$  ms after T1 onset (Broadbent & Broadbent, 1987; Raymond, Shapiro, & Arnell, 1992); a phenomenon known as the attentional blink (AB). There is one salient exception to the AB: When T2 follows T1 immediately at Lag 1, without distractors in between, it is often identified quite well. This exception is called the Lag 1 sparing effect.

Further to the special status of Lag 1, Hommel and Akyürek (2005) showed that although the identity of both targets is often retained, their temporal order is often lost; instead of reporting T1 as the first target and T2 as the second, observers frequently report T1 as the second target and T2 as the first. The frequency of these order errors furthermore varies with the expectations of the observers with regard to stimulus presentation speed (Akyürek, Toffanin, & Hommel, 2008). Hommel and Akyürek interpreted these order errors as a consequence of the temporal integration of the two targets into one event representation, and concluded that temporal integration is likely to play a dominant role at Lag 1 in RSVP. This was confirmed by Akyürek et al. (2012), who presented target stimuli that formed reportable identities not only when viewed individually, but also when combined. They used targets such as “/” and “\” that could be perceptually combined to form an “X,” which itself was then also a possible target identity. In this task observers frequently reported having seen only the integrated percepts at Lag 1 (at the expense of order errors),

confirming the expected effect of temporal integration at this lag. Taken together, these RSVP studies thus suggest that although temporal integration may facilitate visual target identification, it does come at a price—information about the sequence of individual stimuli is lost.

In summary, it seems that in vision, as in audition, two stimuli can also be bound to a single memory trace. Yet, an obvious discrepancy also exists. Whereas in vision temporal integration seems to be associated with a loss of temporal order of the stimuli that are part of the integrated percept, auditory studies, in particular those examining the MMN, suggest that such temporal information is mostly retained. Note that it is entirely possible that this apparent difference between modalities exists as a consequence of the different roles of time in vision and audition: One might argue that the importance of time in audition may render it immune to losses that are incurred in vision, in which spatial information may dominate.

### Current Research

The present study sought to examine the degree to which temporal information and stimulus individuality might be retained in auditory temporal integration, and whether (these aspects of) temporal integration might be modality specific. In particular, the study aimed to provide more definitive evidence of how auditory temporal integration works and to investigate what models are most plausible. To this end, an auditory task similar to RSVP was developed, in which temporal integration of two strictly successive target stimuli was likely. In this rapid serial auditory presentation (RSAP) task (see, e.g., Horváth & Burgyán, 2011; Tremblay, Vachon, & Jones, 2005), the targets were chosen in such a way that both the successive report of individual targets, as well as their combined report, were possible (similar to Akyürek et al., 2012). Targets of the study consisted of pairs of first and second formants (harmonic complexes bandpass filtered at specific frequencies) and the two-formant combined synthetic vowels. In other words, participants were able to report hearing an integrated percept of the sequentially presented formants, which would be equal to a simultaneous presentation thereof (i.e., a two-formant vowel). Reports could thus vary between having heard T1 first and T2 second, T2 and then T1 (order error), or T1 + T2 (integration of first and second formants into two-formant combined synthetic vowels), and any partial version in which either target was missed.

Three versions of the RSAP task were implemented: In Experiment 1, natural differences in formant intensity of the formant pairs as measured from spoken Dutch vowels (Pols, Tromp, & Plomp, 1973) were used for the successive targets. The use of natural differences in intensity means that the first formant (F1) is always of higher intensity than the second formant (F2), resulting in a more natural percept of the two-formant vowels. However, in the visual domain, a large contrast between the physical properties of T1 and T2 can also have an effect on attentional blink and the sparing effect (Experiments 2a and 3, Table 1, Chua, 2005). Therefore, to rule out any additional effects due to differences in intensity (and the resulting loudness), loudness difference was minimized in Experiment 2, where formants of equal loudness, based on the equal-loudness contour (ISO 226:2003; International Organization for Standardization, 2003), were used. As a consequence, the vowels in Experiment 2 sounded less natural, which

also provided a measure of the extent to which natural language familiarity might contribute to integration. Finally, Experiment 3 was performed to investigate the possible effects of the response alternatives that were available to the participants. Because the majority (5/7) of response keys in Experiment 1 and 2 represented vowels, this might have induced a general bias toward reporting vowels. Therefore the number of vowel response keys was reduced (to 1/3) in Experiment 3.

The predictions were as follows. If temporal integration in audition retains temporal coordinates, as suggested by previous work, then the integration of the targets in the present task at short lags (i.e., Lag 1) should result in an increase in the number of correct reports, that is, an escape from the attentional blink. However, neither reports of illusory simultaneous percepts, nor the frequency of order errors should be increased. However, if temporal integration in audition behaves similarly to its visual counterpart, then reports of integrated percepts should be frequent. This would support the idea that temporal integration is a central, modality-unspecific perceptual function.

### Experiment 1

Experiment 1 investigated whether two auditory targets could be integrated and reported as a single integrated percept, using natural intensity differences of the first two formants of naturally spoken Dutch vowels.

### Method

**Participants.** Sixteen (13 female, 3 male) normal hearing (< 20 dB hearing level measured at .25, .5, 1, 2, 4, and 6 kHz) and native Dutch-speaker students of the Psychology Department at the University of Groningen participated in the experiment for course credit. Mean age was 20 years (range 18–23 years). Participants were unaware of the purpose of the experiment. Informed consent was obtained in writing and ethical approval was obtained from the local ethical committee of the Psychology Department.

**Apparatus and stimuli.** The experiment was programmed in Matlab (7.10.0.499 32-bit) using Psychtoolbox (3.0.9; Brainard, 1997; Pelli, 1997) and run under Max OS X (10.5.8) on a Mac Pro equipped with a quad-core Xeon CPU and 8 GB RAM (Apple, Inc., Cupertino, CA). Participants were tested in a sound-isolated booth. Sounds were presented diotically through a Sennheiser HD 600 headphone (Sennheiser Electronic Corporation, Old Lyme, CT), connected to an Echo Audiofire 4 external soundcard (Echo Audio Corp., Santa Barbara, CA) and a Lavry Engineering DA10 digital-to-analog converter (Lavry Engineering, Inc., Rolling Bay, WA). Responses were collected with a standard keyboard.

Target stimuli consisted of first and second formants (F1 and F2), harmonic complexes bandpass filtered (specifics given later and in Table 1) at the formant frequencies, of the five Dutch vowels /a/ (as in *haat*), /i/ (as in *hiet*), /ɪ/ (as in *hit*), /ø/ (as in *heut*), and /y/ (as in *huut*). The synthetic vowel that would result from simultaneous presentation of these formant pairs was also a possible target identity so that the participants could illusorily report a vowel, but it was only rarely an actual target (i.e., on some of the single-target trials). A complex tone with a center frequency of 1 kHz, produced with the same bandpass filter as for the formants, was used as a repeating distractor. One kHz lies between the F1

Table 1  
*Frequencies of F1 and F2 and Deviations of F2 Intensity From F1 Intensity, in dB SPL*

Formant feature	/a/	/i/	/l/	/ø/	/y/
F1 in Hz	795	294	388	443	305
F2 in Hz	1,301	2,208	2,003	1,497	1,730
Deviation of F2 intensity from F1 intensity (dB SPL)	-5.6	-19.5	-17.3	-15.6	-18.1

Note. Frequencies were obtained from Pols, Tromp, and Plomp (1973). F1 = first formant; F2 = second formant.

and F2 values, and therefore fits well with the task that required participants to identify F1 and F2 as low and high tones, respectively. The vowels were specifically chosen, based on the distance in frequency of both formants to the 1 kHz boundary and on the relative distance of the formants between vowels. Larger frequency distances between formants and the 1 kHz boundary were aimed for to increase discriminability between the five vowels.

The formants and distractor stimuli were created by applying an infinite impulse response (IIR) filter (Carlyon, Deeks, Norris, & Butterfield, 2002; Heinrich, Carlyon, Davis, & Johnsrude, 2008; Rabiner & Schafer, 1978) at the desired center frequency (see Table 1; based on Pols et al., 1973) to a harmonic complex of 120 Hz with 100 harmonics and a sampling rate of 44.1 kHz. The filter orders for /a/, /i/, /l/, /ø/, /y/, and the distractor were 6, 10, 4, 6, 10, and 8, respectively, and were empirically chosen based on achieving a balance between creating tone-like stimuli for single targets and vowel-like stimuli once formants were combined. The 3-dB bandwidth of the filter was set at 90 Hz. F1 was presented at 65 dB SPL, but F2 was presented at a lower SPL than the F1 of the same vowel, according to the intensity differences between formants observed in natural speech in the Dutch language (Pols et al., 1973). The vowels, that is, combined formants, as well as the distractors were presented at 65 dB SPL. Figure 1 shows spectrograms, which illustrate the formant stimuli (F1 and F2 of the five vowels) in the lower panels, and an example trial of Lag 3 containing the F1 and F2 of the vowel /a/ together with the surrounding distractors in the upper panel.

**Procedure and design.** Participants were unaware that among the stimuli five different F1s and F2s were used. Instead they were told that the targets consisted of a random low tone (which was an F1), a random high tone (which was an F2) and five vowels. A low tone was defined as any given F1 tone that was lower in frequency than the distractor and a high tone as any given F2 tone higher than the distractor. All seven possible targets were labeled on the numerical keypad, so that participants did not have to memorize which target corresponded to which key on the keyboard.

Participants had to be acquainted to the vowels, learn to distinguish them, and also learn to classify a low and high tone with respect to the distractor. Therefore, in the first session, participants could press any of the labeled keys to hear a stimulus until they felt they could distinguish all five vowels and knew the difference between a low tone, high tone, and distractor. After that session, there was a short training with feedback in which stimuli were presented and participants had to report which of the stimuli they heard. This training session was completed within 15 min on

average. Once participants successfully learned to distinguish the stimuli, there was a short block of practice trials. The only feedback provided was the playback of the sound of the participant's response, so that the participants could compare their response to what was heard in the trial. After that, the real experiment began which consisted of 605 trials with no feedback. A trial consisted of a stream of 18 consecutive items; in this stream there could be either one or two targets, the rest of the items were distractors. On 92.6% of all trials there were two targets. In these two-target trials both formants of a particular vowel were required to be targets (i.e., T1 was F1, T2 was F2, or vice versa). T1 could appear as fifth, sixth, seventh, or eighth item. T2 followed T1 with zero, two, or seven distractors in between (Lag 1, Lag 3, and Lag 8, respectively, and 39.7%, 26.4%, and 26.4% of all trials, respectively). T1 was a solo target in 7.4% of all trials, in which T1 could be a single formant (low tones, 2.47%; high tones, 2.47%) or vowel (2.47%). Each item had a duration of 90 ms, determined in a pilot study, and between the items there was a gap of 10 ms; this gave an SOA of 100 ms. The different conditions are illustrated in Figure 2. Each trial started when the space key was pressed, and participants could take a break between trials. After each trial the participant was asked to enter what they heard as first and second target in the correct order. If no first or second target was heard, they could press the enter key for an empty response. Reporting only one target without entering a second one could thereby be counted as a solo response. The experiment lasted approximately 60 min.

**Data analysis.** First, task performance was examined by analyzing the mean accuracy of T1 and (T2/T1) at Lag 1, 3, and 8. (T2/T1) stands for the accuracy of T2 in cases when T1 was correct. Note that in these analyses a target is only considered correct if both identity and temporal order have been successfully reported. Each analysis consisted of a repeated-measures analysis of variance (ANOVA) with the single variable of lag (1, 3, or 8). In these ANOVAs, when sphericity was not assumed, degrees of freedom were adjusted using the Greenhouse–Geisser epsilon correction. The same analyses were performed for frequency of strict integrations (i.e., only a single integrated response reported) and order reversals (i.e., both targets reported in the incorrect order). Strict integrations and order reversals are cases in which both target identities were preserved; these analyses were therefore conducted relative to the total number of trials on which both target identities were preserved. An example of a strict integration response occurs if T1 is F1 (low tone) and T2 is F2 (high tone) of the vowel /l/ and /l/ is given as a solo response. This indicates that both targets (and thus formants) have been integrated into a single representation of the particular vowel and no second target is perceived. Furthermore, to assess the presence of the attentional blink, a paired samples *t* test was used to compare T2/T1 identification accuracy at Lag 1 to Lag 8. In addition, all analyses were performed on rationalized arcsine transformed scores. The statistical outcomes of these transformed scores are reported when they differed from the analyses on untransformed scores. In all analyses, an alpha level of .05 was used. Each analysis is clarified by line or bar graphs. The line graphs that show strict integrations and order reversals together depict frequencies relative to the total number of trials on which both target identities were preserved, while the bar graphs show absolute report frequencies.

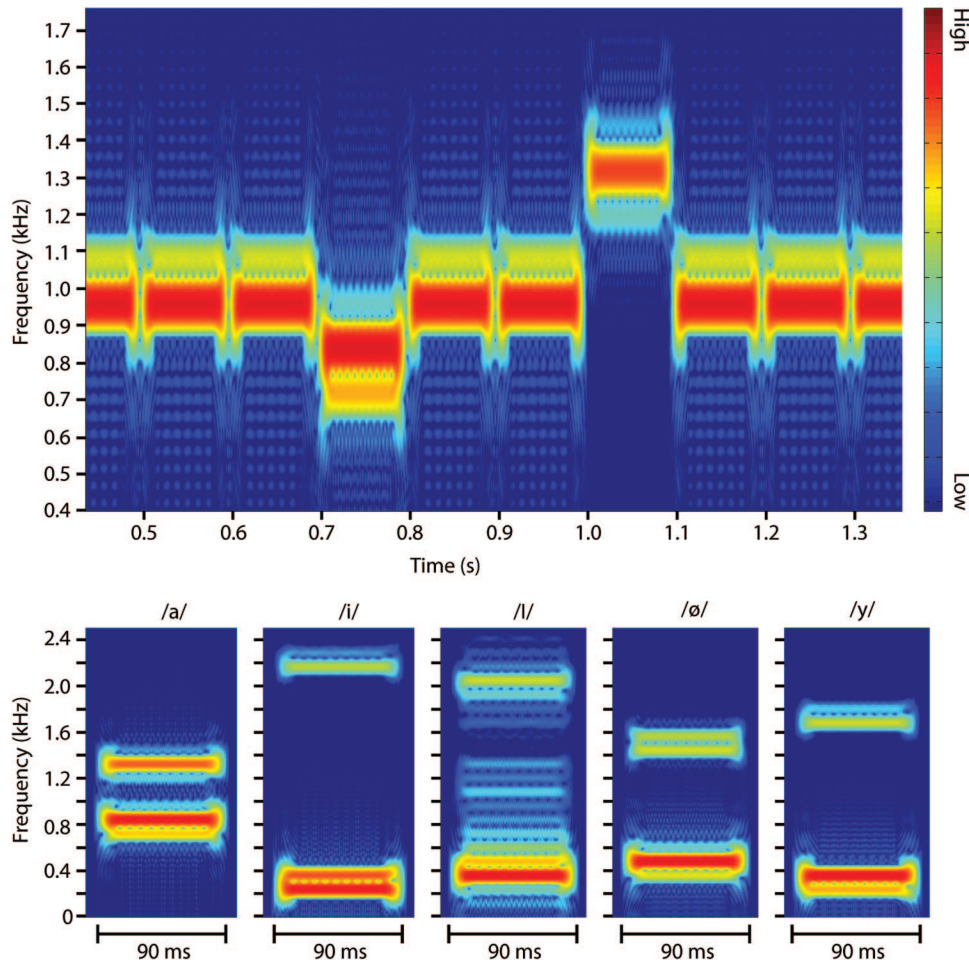


Figure 1. Representation of the stimuli. The top of the figure shows the spectrogram that illustrates a part of a Lag 3 trial. Energy values are represented by different color gradients and range from low (dark blue) to high values (dark red). Complex tones are represented by high concentrations of energy, which last 90 ms and are followed by a silent gap of 10 ms. This example illustrates the midsection of a Lag 3 trial, in which first distractor tones are presented, followed by a low tone (F1 of /a/), then two distractor tones, and a high tone (F2 of /a/) followed by more distractor tones. The five spectrograms at the lower half of the figure illustrate the five two-formant vowels /a/, /i/, /l/, /ø/, and /y/ that were combined by adding the corresponding first and second formants. F1 = first formant; F2 = second formant.

## Results and Discussion

T1 accuracy was strongly affected by lag,  $F(1.4, 20.5) = 17.489$ ,  $MSE = .004$ ,  $p < .001$ . Performance averaged 20.1% at Lag 1, compared to 27.1% at Lag 3, and 31.5% at Lag 8. When report order was ignored performance was 49.2% at Lag 1, 56.6% at Lag 3, and 60.5% at Lag 8. This is illustrated by the left panel of Figure 3.

The accuracy for (T2/T1) was affected by lag,  $F(2, 30) = 5.081$ ,  $MSE = .013$ ,  $p < .015$ . Performance averaged 14.4% at Lag 1, compared to 25% at Lag 3, and 25.7% at Lag 8. A paired samples  $t$  test showed a significant difference between Lag 1 and Lag 8,  $t(15) = -2.989$ ,  $MSE = .038$ ,  $p < .01$ , indicating an early attentional blink (cf. Horváth & Burgyán, 2011; Tremblay et al., 2005). It also indicated, as is often observed in RSAP tasks, that there was no Lag 1 sparing. When report order was ignored,

performance was 67.7% at Lag 1, 69.9% at Lag 3, and 71.5% at Lag 8. This is illustrated by the right panel of Figure 3.

More important, the frequency of strict integrations was strongly affected by lag,  $F(2, 30) = 20.093$ ,  $MSE = .026$ ,  $p < .001$ . Integrations averaged 66.9% at Lag 1, compared to 41.9% at Lag 3, and 31.8% at Lag 8. Order reversals were not affected by Lag,  $F(2, 30) = 2.939$ ,  $MSE = .008$ ,  $p = .068$ . Reversals averaged 8.1% at Lag 1, compared to 15.6% at Lag 3, and 11.5% at Lag 8.<sup>1</sup>

Figure 4 illustrates that the number of strict integrations was higher at Lag 1 compared to later lags. This suggests that two

<sup>1</sup> Analyses on the rationalized arcsine transformed scores show that order reversals were affected by lag,  $F(2, 30) = 4.392$ ,  $MSE = 117.479$ ,  $p < .05$ . Reversals averaged 1.6 rational arcsine units (RAU) at Lag 1, compared to 12.8 RAU at Lag 3, and 8.9 RAU at Lag 8.

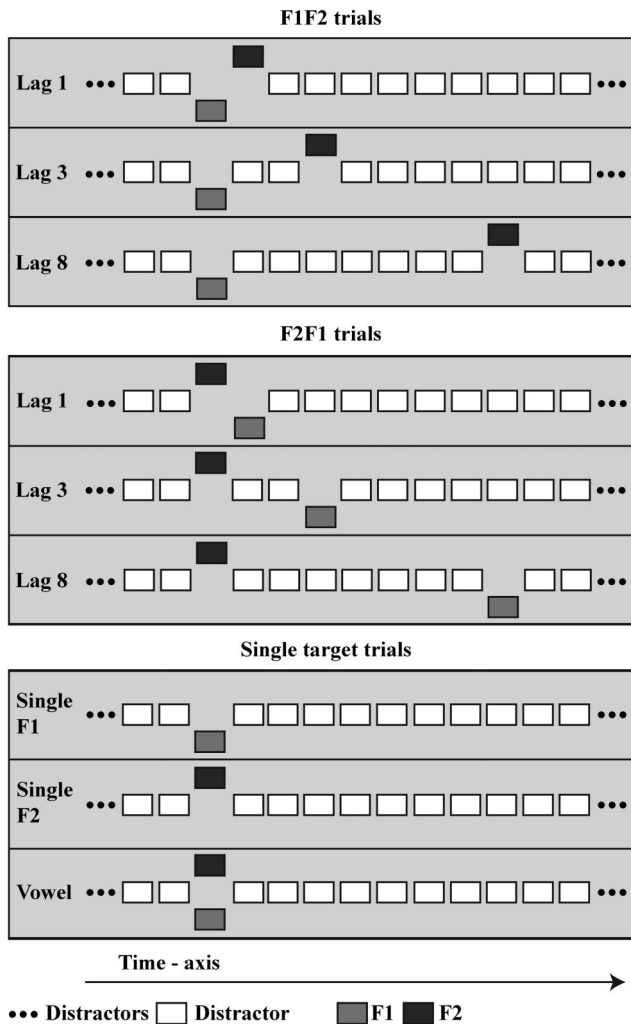


Figure 2. Schematic representation of the different conditions. The number that accompanies the lag indicates the temporal delay between the first and second target: for example, Lag 3 means that T2 lags T1 as the third successive stimulus with two distractors in between. The height of the items indicates the relative frequency differences, for example, F1s have lower frequency than distractors, which in turn have lower frequency than F2s. Targets as well as distractors lasted 90 ms, followed by a silent gap of 10 ms. F1 = first formant; F2 = second formant; T2 = second target; T1 = first target.

distinct auditory stimuli that succeed each other in a short interval, without actually overlapping or being physically continuous, can indeed be temporally integrated in such a way that a meaningful percept is constructed. The report of such integrated percepts implies that its constituent tones were perceived as if they were simultaneous; a complete loss of order information similar to that observed in visual temporal integration (Akyürek et al., 2012). In this context it is important to note that singular integrations (i.e., without entering a second response) were reported despite deliberate biases in the task toward the report of two individual tones, which were by far the most frequent stimuli, and the most frequent type of trial. Indeed, at later lags, increased reports of the two individual targets were

observed. At these lags the succession between targets is too slow and together with the presence of intervening distractors, makes integration unlikely.

## Experiment 2

Experiment 2 was designed to eliminate potential effects of intensity contrast between F1 and F2, as well as possible resultant language familiarity effects, as discussed earlier, by presenting all stimuli at the same loudness.

## Method

**Participants.** Sixteen (12 female, 4 male) new participants were included using the same procedures and criteria as in Experiment 1. The mean age was 20 years with a range of 18 to 25 years.

**Apparatus and stimuli.** The experimental setup and stimuli were the same as for Experiment 1. The only difference was that the relative intensity differences between formants from Table 1 were not used. Instead, each stimulus was presented at the same loudness, determined using the equal-loudness contour (ISO 226: 2003; International Organization for Standardization, 2003). This contour gives estimates of what intensity level in dB SPL is needed for a stimulus to sound subjectively equally loud as a stimulus of 1 kHz at a particular loudness level in phons. Table 2 shows the values in dB SPL that were obtained by the calculations using the equal-loudness contours. All F2s were adjusted to these values. Vowels were presented at the average SPL of both corresponding formants.

**Procedure and design.** The procedure and design were the same as in the previous experiment.

## Results and Discussion

T1 accuracy was not affected by lag,  $F(1.3, 19.1) = 3.441$ ,  $MSE = .007$ ,  $p = .071$ . Performance averaged at 26.6% at Lag 1, compared to 31.3% at Lag 3, and 32.2% at Lag 8. When report order was ignored performance was 54.9% at Lag 1, 61% at Lag 3, and 61.1% at Lag 8. This is illustrated in the left panel of Figure 5.

Accuracy for (T2|T1) was strongly affected by lag,  $F(2, 30) = 10.006$ ,  $MSE = .014$ ,  $p < .001$ . Performance averaged at 22.1% at Lag 1, compared to 35.6% at Lag 3, and 39.8% at Lag 8. A paired samples  $t$  test showed a significant difference between Lag 1 and Lag 8,  $t(15) = -3.896$ ,  $MSE = .045$ ,  $p < .001$ , indicating the expected early attentional blink, similar to the previous experiment, despite using equal loudness for all stimuli. When report order was ignored, performance was 66.5% at Lag 1, 71.7% at Lag 3, and 73.8% at Lag 8. This is illustrated in the right panel of Figure 5.

The frequency of strict integration was again strongly affected by lag,  $F(1.3, 19.4) = 23.280$ ,  $MSE = .058$ ,  $p < .001$ . Integrations averaged 60.9% at Lag 1, compared to 20.9% at Lag 3, and 19.7% at Lag 8. Order reversals were not affected by lag,  $F(1.3, 19) = 2.297$ ,  $MSE = .036$ ,  $p = .142$ . Reversals averaged 7.2% at Lag 1, compared to 17.8% at Lag 3, and 8.6% at Lag 8.<sup>2</sup>

<sup>2</sup> Analyses on the rationalized arcsine transformed scores show that order reversals were affected by lag,  $F(2, 30) = 3.472$ ,  $MSE = 222.842$ ,  $p < .05$ . Reversals averaged 1.7 RAU at Lag 1, compared to 14.3 RAU at Lag 3, and 2.8 RAU at Lag 8.

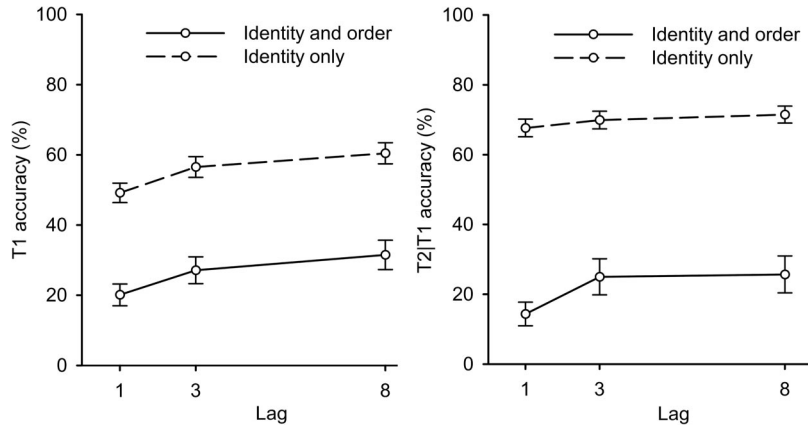


Figure 3. Experiment 1: The left panel shows task performance on T1 in percentage correct, plotted over lag (T2 being first, third, or eighth stimulus after T1). Error bars represent  $\pm 1$  standard error of the mean. The right panel shows T2 performance given that T1 was correctly reported (T2/T1) in percentage correct plotted over lag. Dashed lines represent identification accuracy if report order is ignored (relaxed accuracy criterion). T2 = second target; T1 = first target.

Figure 6 also illustrates how, similar to Experiment 1, the relatively high number of integrations at Lag 1 stands in contrast to that at the longer lags. The number of order reversals was not affected by lag and seemed, similar to Experiment 1, unrelated to integration frequency. Overall, Experiment 2 replicated the results of Experiment 1. Thus, it can be concluded that temporal integration was not the result of the loudness differences between the stimuli that were used in Experiment 1, and also was unlikely to result from the degree of familiarity with the vowels used in the task.

### Experiment 3

Experiment 3 was conducted to eliminate a possible response bias toward the report of vowels by reducing the number of vowel

response alternatives. To this end, the number of vowel stimuli (and consequently the respective F1s and F2s) was reduced from five to three. Next to these three vowel response alternatives, participants now had the opportunity to identify the six remaining tones (rather than just classify as high or low), which made up the majority of the response alternatives (6/9).

### Method

**Participants.** Fifteen (9 female, 6 male) normal hearing (< 20 dB hearing level measured at .25, .5, 1, 2, 4, and 6 kHz) and native Dutch-speaker students of the Psychology Department at the University of Groningen participated in the experiment following the same procedure as in Experiment 1. Mean age was 21 years (range 20–23 years).

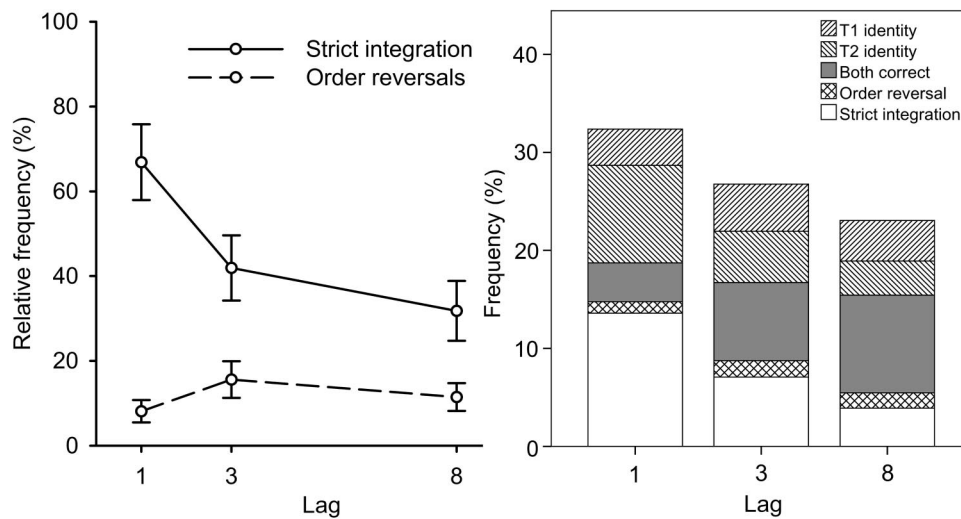


Figure 4. Experiment 1: The left panel shows the relative frequency of strict integrations and order reversals plotted over lag, as a percentage of the total number of responses in which both target identities were preserved. The right panel shows the distribution of responses for each lag, as a percentage of the total number of responses. T2 = second target; T1 = first target.

Table 2  
Sound Pressure Levels Calculated With Equal Loudness Contours

Formant feature	Distractor	/a/	/i/	/ʌ/	/ø/	/y/
F1 center frequency (Hz)	1,000	795	294	388	443	305
F1 intensity (dB SPL)	65	64.8	70.3	68.2	67.4	70
F2 center frequency (Hz)	1,000	1,301	2,208	2,003	1,497	1,730
F2 intensity (dB SPL)	65	67.7	63.6	65.1	68.5	67.6

Note. Distractor = 65 Phon at 1 kHz; F1 = first formant; F2 = second formant.

**Apparatus and stimuli.** Apparatus and stimuli were similar to that of Experiment 2, except that only three Dutch vowels were used as stimuli: /a/ (as in *haat*), /i/ (as in *hiet*), and /ø/ (as in *heut*).

**Procedure and design.** The task differed from the previous two experiments such that when a tone was heard as a target, the participants not only had to classify it as low or high with respect to the filler tone, but additionally they had to identify the correct tone among three different low and three different high tone options. Thus, the response alternatives were three vowels, three low, and three high tones. This increased task difficulty, but more important removed any response bias toward vowels, as the vowel response distribution was three out of nine choices, instead of five out of seven as in the previous experiments.

The task consisted of 549 trials with no feedback. On 91.8% of all trials there were two targets. T2 followed T1 with zero, two, or seven distractors in between (Lag 1, Lag 3, and Lag 8, respectively, and 39.3%, 26.2%, and 26.2% of all trials, respectively). T1 was a solo target in 8.2% of all trials in which T1 could be a single formant or vowel; each of the nine response alternatives was a solo target in 0.91% of all trials. The experiment lasted approximately 60 min.

## Results and Discussion

T1 accuracy was strongly affected by lag,  $F(2, 28) = 12.271$ ,  $MSE = .003$ ,  $p < .001$ . Performance averaged 26% at Lag 1,

compared to 32.3% at Lag 3, and 35.9% at Lag 8. When report order was ignored performance was 37.5% at Lag 1, 38.3% at Lag 3, and 42.3% at Lag 8. This is illustrated by the left panel of Figure 7.

The accuracy for (T2|T1) was affected by lag,  $F(2, 28) = 3.562$ ,  $MSE = .011$ ,  $p < .05$ . Performance averaged 28.9% at Lag 1, compared to 33.9% at Lag 3, and 39.2% at Lag 8. A paired samples  $t$  test showed a significant difference between Lag 1 and Lag 8,  $t(14) = -2.550$ ,  $MSE = .040$ ,  $p < .05$ , again indicating an early attentional blink. When report order was ignored, performance was 45.4% at Lag 1, 37% at Lag 3, and 40.8% at Lag 8, as shown in the right panel of Figure 7.

More important, the frequency of strict integrations was strongly affected by lag,  $F(1.1, 15.4) = 12.208$ ,  $MSE = .082$ ,  $p < .005$ . Integrations averaged 35.3% at Lag 1, compared to 5% at Lag 3, and 0% at Lag 8. Order reversals were not affected by lag,  $F(2, 28) = 0.483$ ,  $MSE = .009$ ,  $p = .622$ . Reversals averaged 10.5% at Lag 1, compared to 9.2% at Lag 3, and 7% at Lag 8.

Figure 8 shows that despite the fact that Lag 1 was not completely dominated by strict integrations, as was the case with the previous two experiments, the number of strict integrations was still relatively high at Lag 1. Indeed, strict integrations were almost solely present at Lag 1. Integration of targets at longer intervals (Lag 3 and 8), and with multiple intervening distractors, was not necessarily predicted, and so the absence of integration reports at

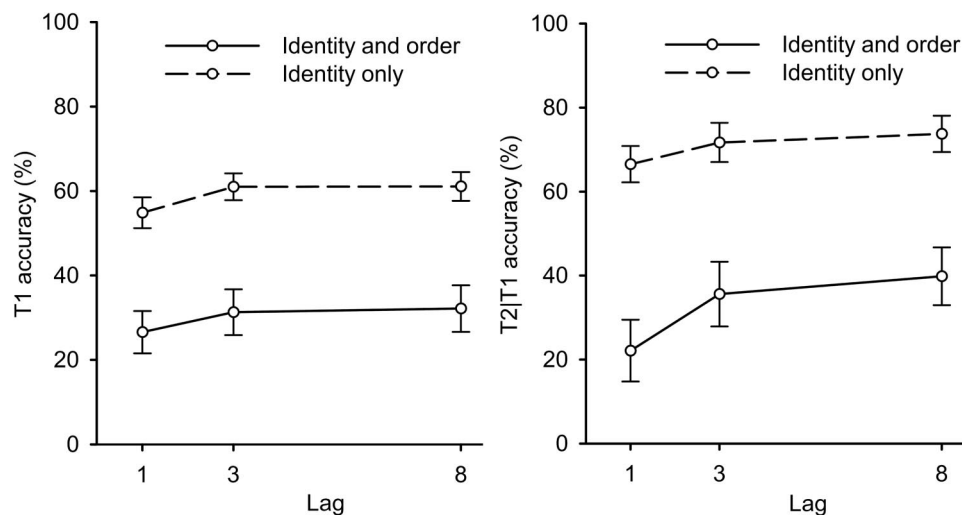


Figure 5. Experiment 2: The left panel shows T1 task performance for each lag. Error bars represent  $\pm 1$  standard error of the mean. The right panel shows (T2|T1) performance for each lag. Dashed lines represent identification accuracy if report order is ignored. T2 = second target; T1 = first target.



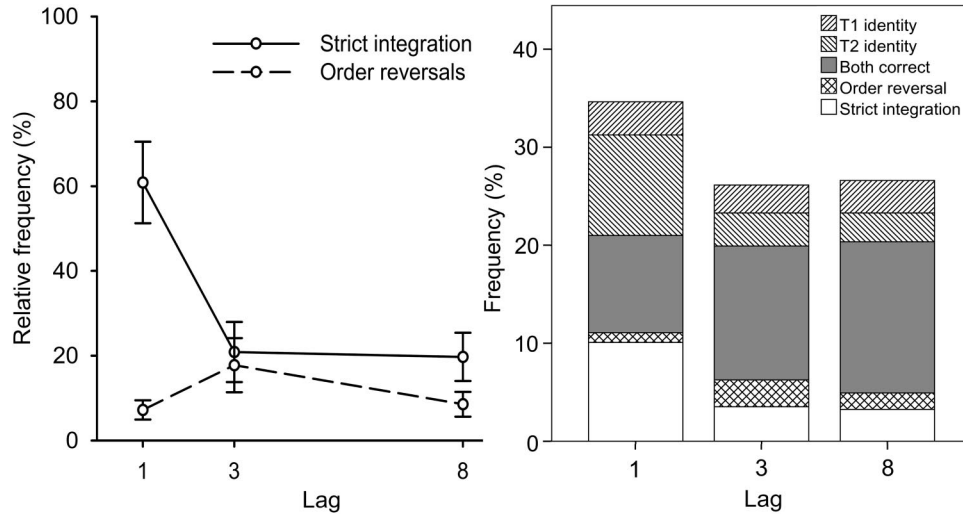


Figure 6. Experiment 2: The left panel shows the relative frequency of strict integrations and order reversals for each lag, as a percentage of the total number of responses in which both target identities were preserved. The right panel shows the distribution of responses for each lag, as a percentage of the total number of responses. T1 = first target; T2 = second target.

these longer lags was in line with expectations. The “baseline” frequency of integration reports at these longer lags in the previous experiments is thus indeed likely to have resulted from a response bias toward vowels, which the present experiment removed. More important, however, at Lag 1, where integration is expected, the number of integrations remained substantial. The frequency of order reversals, on the other hand, still did not change across lags.

**General Discussion**

The present study investigated whether two rapidly following auditory stimuli can be integrated and perceived as if they were presented simultaneously, resulting in a unitary integrated percept,

similar to what is commonly observed in the visual domain. This was confirmed in three versions of an RSAP task. Participants indeed frequently only reported an integrated percept of a synthetic vowel at Lag 1, while such reports were rare at longer lags, consistent with the idea of temporal integration. The perception of a single synthetic vowel when two complex tones were presented nonsimultaneously (at Lag 1) also shows that temporal integration is much more complex than simple energy summation, an interpretation previously given by some authors (Pedersen & Elberling, 1972; Pedersen & Salomon, 1977; Zwislocki, 1969). In converse to that, the combined results of the present experiments also suggested that integration does not rely heavily on high-level

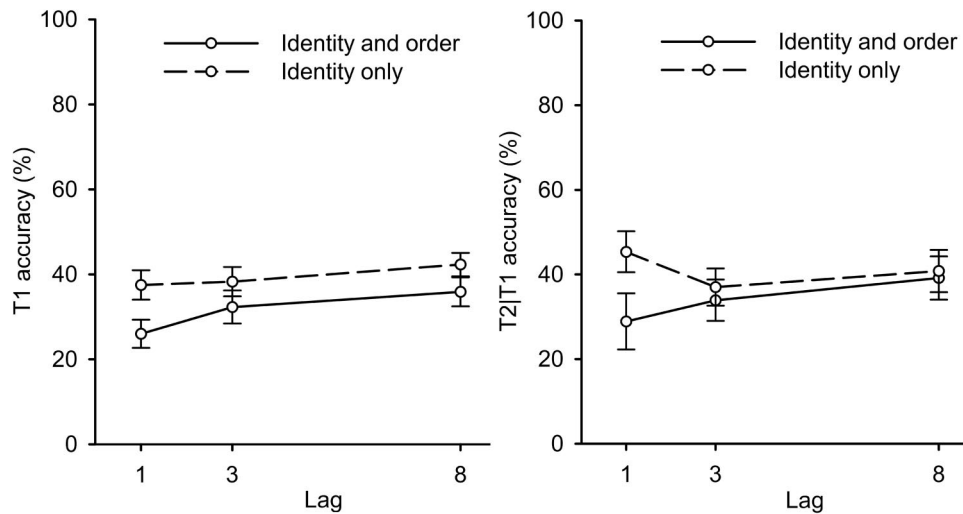


Figure 7. Experiment 3: The left panel shows T1 task performance for each lag. Error bars represent  $\pm 1$  standard error of the mean. The right panel shows (T2/T1) performance for each lag. Dashed lines represent identification accuracy if report order is ignored. T2 = second target; T1 = first target.

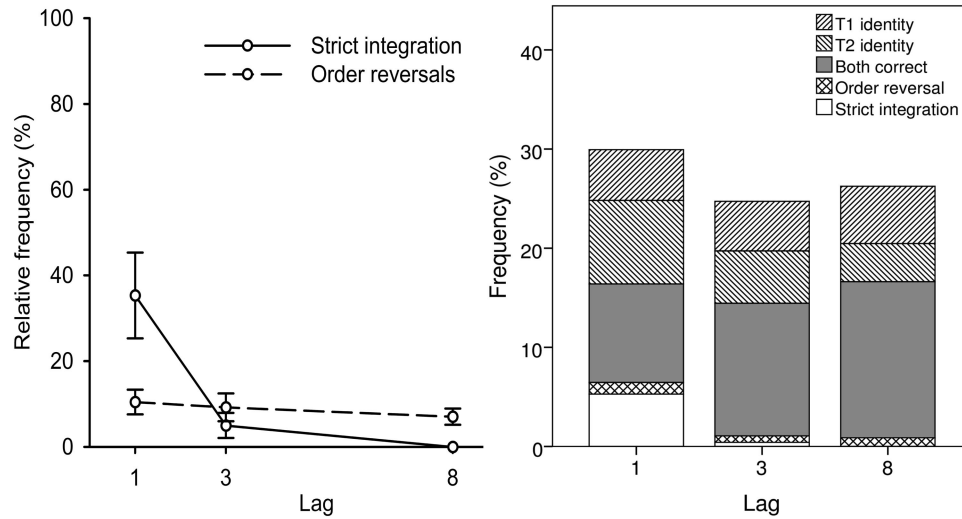


Figure 8. Experiment 3: The left panel shows the relative frequency of strict integrations and order reversals for each lag, as a percentage of the total number of responses in which both target identities were preserved. The right panel shows the distribution of responses for each lag, as a percentage of the total number of responses. T1 = first target; T2 = second target.

(linguistic) knowledge either: Integration was as frequent with natural formant intensity ratios as it was without.

The current findings are overall most compatible with more comprehensive accounts of temporal integration, such as discussed by Moore (2003) or Näätänen and Winkler (1999), except for the fact that they hypothesized that acoustic information is integrated and placed on temporal coordinates while the present data show that temporal information is often lost. When identity information of both targets was retained, participants frequently reported to hear only the integrated percept of a synthetic vowel, which was the correct assembly, but also the temporal merger of both target formants, instead of reporting both targets in the correct or incorrect order, despite the inherently high temporal resolution of the auditory system (Eddins & Green, 1995).

It is interesting that in cases when two targets were heard, order information did not seem to suffer from the temporal proximity of the targets. This contrasts with the findings obtained in visual tasks, which do show an increase in order errors at Lag 1, even if their frequency is relatively low overall (Akyürek et al., 2012). If anything, order errors were reduced at Lag 1, at least in Experiment 1 and 2, although this might also be a consequence of a reduced ability to separate the targets in the first place. At subsequent lags, when reports of integrated percepts decreased, there was a proportional increase in fully correct responses, while order reversals remained infrequent, but relatively constant across lags. When temporal integration does not occur, it thus seems the auditory system does keep close track of stimulus order.

### Relationship to Previous Studies on Tone Perception

Findings from MMN studies may at first glance appear to contrast with the present results. However, although deviance detection in MMN studies seems to suggest that temporal order is retained within integrated percepts, this may not be a necessary assumption. Grouping pairs (or more) of stimuli together in one

percept and dissociating it from other tones that occur after longer delays only requires that the integrated percept is perceived in time in reference to other percepts. It does not necessarily require that its constituent parts also are ordered correctly in time.

Some findings of Tervaniemi et al. (1994) provide some further support for this view. In their study, pairs of two different tones were presented in series, separated by silent gaps. During this continuous stream, when the second tone of a pair was omitted, an MMN was elicited. Thus, one might conclude that each tone pair was regarded as a unitary event and that the listener expected to perceive the first and second tone of the pair in order, as the definition says that integrated stimuli are placed on temporal coordinates (Näätänen & Winkler, 1999). Yet, this account seems inconsistent with the fact that no statistically significant MMN was elicited when Tervaniemi et al. reversed the first and second tone of the pair, instead of omitting the second tone. This suggests that a deviant, order-reversed tonal pair is not regarded as a deviance from the norm by the auditory system (per se). Although the absence of an MMN as such in this study may not be fully conclusive, the observed nondeviance of an order-reversed tone pair does suggest that order information within the perceptual event might have been missing.

Findings from another study conducted by Ciocca and Darwin (1999), focusing on pitch perception, might also support the idea that integrated auditory stimuli are not placed on temporal coordinates. In this study, nonsimultaneous mistuned sound components presented temporally close to a target sound changed its perceived pitch. However, by themselves, these results can also be accounted for by assuming that pitch processes work from samples in STM (as proposed in the multiple-looks model by Viemeister, 1996), creating a virtual pitch without losing the individuality and temporal order of the stored samples.

The current findings are, however, largely incompatible with the multiple-looks model, as it assumes that there is no true long-term

integration of the kind that we report here (Viemeister, 1996). The multiple-looks model states that long-term temporal integration (of the magnitude presently obtained) can only be achieved by computations made on short-term samples ( $\pm 3$  ms) in STM. The present data clearly show that the information available for report (i.e., in STM) does not seem to consist of individual samples; instead, these appear to have been lost or irreversibly overlaid by an integrated percept—otherwise there would be no reason for not reporting the individual targets in the present task. Recall that participants were in principle expecting to be able to report two targets, not just one, and that the integrated percept was in fact much rarer than the formants, thus making the latter unappealing as a response choice from a strategic perspective.

To accommodate the present results, the multiple-looks model could possibly be modified by allowing the computations that are assumed to apply to multiple samples in STM to act as a kind of long-term temporal integration window (of a few hundred ms), which assembles the samples into a single acoustic percept at the expense of the individuality of the samples. However, this would seem to go against one of the principal tenets of the model, namely that integration across longer intervals does not take place (Viemeister, 1996).

### Relationship to the Continuity Illusion and Phonemic Restoration

One might suspect that auditory temporal integration is related to the continuity illusion, which is the perception of a discontinuous, interrupted signal as being continuous when the gaps are filled by loud noise (Başkent, Eiler, & Edwards, 2009; Carlyon et al., 2002; Heinrich et al., 2008; Warren, Obusek, & Ackroff, 1972). The illusion is strongest when temporal and spectral components from the noise are matched to those from the sound signal (Bregman, 1994; Riecke, van Opstal, & Formisano, 2008; Warren, 1999). During the continuity illusion, separate (interrupted) stimuli are perceived as one coherent signal, which is similar to temporal integration, as the current research showed that people can perceive two stimuli (separated by a temporal gap) as a single integrated entity. Furthermore, in a previous study, MMN latency data indicated that the processes underlying the continuity illusion are active within a period of 200 ms after the onset of the noise-filled gap, an interval that is comparable to that of temporal integration (Michey et al., 2003).

There is, however, no evidence that during the continuity illusion the continuously perceived entity is being integrated into a single, overlaid entity. To wit, tone sweeps gliding upward or downward in frequency can be perceived as continuous when they are interrupted and their silent gaps filled with noise (Ciocca & Bregman, 1987). In other words, people perceive the tone sweep to continue during the noise with a similar upward or downward trend as before the interruption occurred. Would there be a true integration, then a compound of tones with different frequencies might be a more likely percept instead. Furthermore, the continuity illusion for steady-state tones can still occur when the intervening noise is up to 2,000 ms long (Riecke et al., 2008). Such a duration lies outside the scope of the temporal window of integration. Last, as these examples also show, for the continuity illusion to work, a filler stimulus is needed, such as noise, to bridge the silent gap.

Temporal integration requires no such masker, which, in fact, might even impair integration.

Because of its more linguistic nature, a special case of the continuity illusion may be particularly relevant to temporal integration as presently tested: Phonemic restoration, which is the ability to perceptually restore and enhance intelligibility of interrupted, degraded speech (Başkent, 2012; Başkent et al., 2009; Warren & Sherman, 1974). Phonemic restoration is commonly observed with interrupted speech that has comparable speech and silent/noise intervals to that of temporal integration, but phonemic restoration is clearly more complicated, as it is an interaction between top-down and bottom-up factors, including expectations, linguistic skills, situational, and semantic context, Gestalt rules, as well as spectral and temporal cues from the speech (Bashford, Riener, & Warren, 1992; Başkent, 2012; Davis & Johnsrude, 2007; Samuel, 1981; Stenfelt & Rönnberg, 2009). Nonetheless, in the current task, to identify the correct vowel after both formants are integrated, some knowledge of the vowels from the response alternatives was applied, and a possible role of attentional selection (or top-down control) seems feasible also.

There is indeed prior evidence for common ground between temporal integration and restoration of degraded speech. Using a speech restoration task, Saberi and Perrott (1999) showed that speech intelligibility was almost perfect when speech segments of 50 ms were reversed in time, and only decreased when segments of 100 ms were reversed. Nonetheless, when participants repeatedly listened to stimuli from the latter condition, they reported that the words gradually became clearer and easier to understand, and they eventually reported actually hearing the words. While these segments were reversed in nature (i.e., temporally distorted), it seemed there was still enough information for the auditory system to reconstruct meaningful objects. Temporal coordinates might thus not be fully fixed and may be reordered or reinterpreted if needed.

Although the results of Saberi and Perrott (1999) are intriguing, it seems likely that perceptual organization works differently for meaningful speech units, especially for context-rich sentences (Clarke, Gaudrain, Chatterjee, & Başkent, 2013), than with simpler auditory stimuli. Considering that some practice was needed in the study of Saberi and Perrott, it seems likely that the perceptual reconstruction involved both a reutilization of other speech cues that were not distorted, as well as top-down processes, such as expectancies, linguistic skills, and vocabulary, to correctly interpret the distorted speech signal (Bashford et al., 1992; Başkent, 2012; Davis & Johnsrude, 2007; Samuel, 1981). In other words, perhaps their results would have been different if, instead of highly redundant speech, simpler speech materials were used, such as vowels, syllables, or words without context. The present results address such doubts to an extent: Temporal integration as measured in the present task seems to confirm that temporal coordinates may not always play an important role in the perception of brief events.

### Conclusions

When successive, broadly compatible tones are perceived across an interval of up to 200 ms, temporal integration of these stimuli frequently may give rise to a unified percept that consists of featural properties of the individual tones, but which (strongly)

diminishes their individuality and temporal properties. Thus, temporal integration in the auditory domain is similar to that observed in vision, supporting the view that temporal integration may be a general, amodal perceptual processing function in the human brain.

## References

- Akyürek, E. G., Eshuis, S. A. H., Nieuwenstein, M. R., Saija, J. D., Başkent, D., & Hommel, B. (2012). Temporal target integration underlies performance at lag 1 in the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance*, *38*, 1448–1464. doi:10.1037/a0027610
- Akyürek, E. G., Schubö, A., & Hommel, B. (2010). Fast temporal event integration in the visual domain demonstrated by event-related potentials. *Psychophysiology*, *47*, 512–522. doi:10.1111/j.1469-8986.2010.00962.x
- Akyürek, E. G., Toffanin, P., & Hommel, B. (2008). Adaptive control of event integration. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 569–577. doi:10.1037/0096-1523.34.3.569
- Bashford, J. A., Rienner, K. R., & Warren, R. M. (1992). Increasing the intelligibility of speech through multiple phonemic restorations. *Perception & Psychophysics*, *51*, 211–217. doi:10.3758/BF03212247
- Başkent, D. (2012). Effect of speech degradation on top-down repair: Phonemic restoration with simulations of cochlear implants and combined electric–acoustic stimulation. *The Journal of the Association for Research in Otolaryngology*, *13*, 683–692. doi:10.1007/s10162-012-0334-3
- Başkent, D., Eiler, C. L., & Edwards, B. (2009). Effects of envelope discontinuities on perceptual restoration of amplitude-compressed speech. *The Journal of the Acoustical Society of America*, *125*, 3995–4005. doi:10.1121/1.3125329
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436. doi:10.1163/156856897X00357
- Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press. doi:10.1121/1.408434
- Broadbent, D. E., & Broadbent, M. H. (1987). From detection to identification: Response to multiple targets in rapid serial visual presentation. *Perception & Psychophysics*, *42*, 105–113. doi:10.3758/BF03210498
- Carlyon, R. P., Deeks, J., Norris, D., & Butterfield, S. (2002). The continuity illusion and vowel identification. *Acta Acustica United With Acustica*, *88*, 408–415.
- Chua, F. K. (2005). The effect of target contrast on the attentional blink. *Perception & Psychophysics*, *67*, 770–788. doi:10.3758/BF03193532
- Ciocca, V., & Bregman, A. S. (1987). Perceived continuity of gliding and steady-state tones through interrupting noise. *Perception & Psychophysics*, *42*, 476–484. doi:10.3758/BF03209755
- Ciocca, V., & Darwin, C. J. (1999). The integration of nonsimultaneous frequency components into a single virtual pitch. *The Journal of the Acoustical Society of America*, *105*, 2421–2430. doi:10.1121/1.426847
- Clarke, J., Gaudrain, E., Chatterjee, M., & Başkent, D. (2013). *Perceptual continuity and top-down restoration of speech*. Unpublished manuscript.
- Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hearing Research*, *229*, 132–147. doi:10.1016/j.heares.2007.01.014
- Eddins, D. A., & Green, D. M. (1995). Temporal integration and temporal resolution. In B. C. J. Moore (Ed.), *Hearing* (pp. 207–242). San Diego, CA: Academic Press.
- Heinrich, A., Carlyon, R. P., Davis, M. H., & Johnsrude, I. S. (2008). Illusory vowels resulting from perceptual continuity: A functional magnetic resonance imaging study. *Journal of Cognitive Neuroscience*, *20*, 1737–1752. doi:10.1162/jocn.2008.20069
- Hogben, J. H., & Di Lollo, V. (1974). Perceptual integration and perceptual segregation of brief visual stimuli. *Vision Research*, *14*, 1059–1069. doi:10.1016/0042-6989(74)90202-8
- Hommel, B., & Akyürek, E. G. (2005). Lag-1 sparing in the attentional blink: Benefits and costs of integrating two events into a single episode. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *58*, 1415–1433. doi:10.1080/02724980443000647
- Horváth, J., & Burgián, A. (2011). Distraction and the auditory attentional blink. *Attention, Perception, & Psychophysics*, *73*, 695–701. doi:10.3758/s13414-010-0077-3
- Hughes, J. W. (1946). The threshold of audition for short periods of stimulation. *Proceedings of the Royal Society of London Series B, Biological Sciences*, *133*(873), 486–490. doi:10.1098/rspb.1946.0026
- International Organization for Standardization. (2003). *ISO 226:2003 Acoustics—Normal equal loudness-level contours*. Geneva, Switzerland: Author. Available at [http://www.iso.org/iso/home/store/catalogue\\_tc/catalogue\\_detail.htm?csnumber=34222](http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=34222)
- Jeffress, L. A. (1967). Stimulus-oriented approach to detection re-examined. *The Journal of the Acoustical Society of America*, *41*, 480–488. doi:10.1121/1.1910358
- Matlab 7.10.0.499 32-bit, The MathWorks, Inc., Natick, MA
- Micheyl, C., Carlyon, R. P., Shtyrov, Y., Hauk, O., Dodson, T., & Pullvermüller, F. (2003). The neurophysiological basis of the auditory continuity illusion: A mismatch negativity study. *Journal of Cognitive Neuroscience*, *15*, 747–758. doi:10.1162/jocn.2003.15.5.747
- Moore, B. C. J. (2003). Temporal integration and context effects in hearing. *Journal of Phonetics*, *31*, 563–574. doi:10.1016/S0095-4470(03)00011-1
- Munson, W. A. (1947). The growth of auditory sensation. *The Journal of the Acoustical Society of America*, *19*, 584–591. doi:10.1121/1.1916525
- Näätänen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral and Brain Sciences*, *13*, 201–233. doi:10.1017/S0140525X00078407
- Näätänen, R. (1995). The mismatch negativity: A powerful tool for cognitive neuroscience. *Ear and Hearing*, *16*, 6–18. doi:10.1097/00003446-199502000-00002
- Näätänen, R., Kujala, T., & Winkler, I. (2011). Auditory processing that leads to conscious perception: A unique window to central auditory processing opened by the mismatch negativity and related responses. *Psychophysiology*, *48*, 4–22. doi:10.1111/j.1469-8986.2010.01114.x
- Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, *118*, 2544–2590. doi:10.1016/j.clinph.2007.04.026
- Näätänen, R., & Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin*, *125*, 826–859. doi:10.1037/0033-2909.125.6.826
- Pedersen, C. B., & Elberling, C. (1972). Temporal integration of acoustic energy in normal hearing persons. *Acta Oto-Laryngologica*, *74*, 398–405. doi:10.3109/00016487209128469
- Pedersen, C. B., & Salomon, G. (1977). Temporal integration of acoustic energy. *Acta Oto-Laryngologica*, *83*, 417–423. doi:10.3109/00016487709128866
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442. doi:10.1163/156856897X00366
- Plomp, R., & Bouman, M. A. (1959). Relation between hearing threshold and duration for tone pulses. *The Journal of the Acoustical Society of America*, *31*, 749–758. doi:10.1121/1.1907781
- Pols, L. C. W., Tromp, H. R. C., & Plomp, R. (1973). Frequency analysis of Dutch vowels from 50 male speakers. *The Journal of the Acoustical Society of America*, *53*, 1093–1101. doi:10.1121/1.1913429

- Rabiner, L. R., & Schafer, R. W. (1978). *Digital processing of speech signals* (Vol. 100). Englewood Cliffs, NJ: Prentice-Hall. Retrieved from <http://sibese.sibdi.ucr.ac.cr/dspace/handle/2327/292672>
- Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 849–860. doi:10.1037/0096-1523.18.3.849
- Riecke, L., van Opstal, A. J., & Formisano, E. (2008). The auditory continuity illusion: A parametric investigation and filter model. *Perception & Psychophysics*, *70*, 1–12. doi:10.3758/PP.70.1.1
- Saberí, K., & Perrott, D. R. (1999). Cognitive restoration of reversed speech. *Nature*, *398*(6730), 760. doi:10.1038/19652
- Samuel, A. G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General*, *110*, 474–494. doi:10.1037/0096-3445.110.4.474
- Stenfelt, S., & Rönnberg, J. (2009). The signal-cognition interface: Interactions between degraded auditory signals and cognitive processes. *Scandinavian Journal of Psychology*, *50*(5), 385–393. doi:10.1111/j.1467-9450.2009.00748.x
- Sussman, E., Winkler, I., Ritter, W., Alho, K., & Näätänen, R. (1999). Temporal integration of auditory stimulus deviance as reflected by the mismatch negativity. *Neuroscience Letters*, *264*(1–3), 161–164. doi:10.1016/S0304-3940(99)00214-1
- Tervaniemi, M., Saarinen, J., Paavilainen, P., Danilova, N., & Näätänen, R. (1994). Temporal integration of auditory information in sensory memory as reflected by the mismatch negativity. *Biological Psychology*, *38*(2–3), 157–167. doi:10.1016/0301-0511(94)90036-1
- Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology*, *6*, 171–178. doi:10.1016/S0959-4388(96)80070-5
- Tremblay, S., Vachon, F., & Jones, D. M. (2005). Attentional and perceptual sources of the auditory attentional blink. *Perception & Psychophysics*, *67*, 195–208. doi:10.3758/BF03206484
- Viemeister, N. (1996). Auditory temporal integration: What is being accumulated? *Current Directions in Psychological Science*, *5*, 28–32. doi:10.1111/1467-8721.ep10772699
- Viemeister, N. F., & Wakefield, G. H. (1991). Temporal integration and multiple looks. *The Journal of the Acoustical Society of America*, *90*, 858–865. doi:10.1121/1.401953
- Warren, R. M. (1999). *Auditory perception: A new analysis and synthesis* (Vol. xiv). New York, NY: Cambridge University Press.
- Warren, R. M., Obusek, C. J., & Ackroff, J. M. (1972). Auditory induction: Perceptual synthesis of absent sounds. *Science*, *176*(4039), 1149–1151. doi:10.1126/science.176.4039.1149
- Warren, R. M., & Sherman, G. L. (1974). Phonemic restorations based on subsequent context. *Perception & Psychophysics*, *16*, 150–156. doi:10.3758/BF03203268
- Yabe, H., Tervaniemi, M., Sinkkonen, J., Huotilainen, M., Ilmoniemi, R. J., & Näätänen, R. (1998). Temporal window of integration of auditory information in the human brain. *Psychophysiology*, *35*, 615–619. doi:10.1017/S0048577298000183
- Yu, L., Yabe, H., Shiga, T., Nozaki, M., Ohshima, H., Itagaki, S., . . . Niwa, S. (2011). Only a stimulus onset might initiate the temporal window of integration. In *2011 IEEE/ICME International Conference on Complex Medical Engineering* (pp. 246–247). Heilongjiang, China: IEEE. doi:10.1109/ICCME.2011.5876743
- Zwislocki, J. J. (1960). Theory of temporal auditory summation. *The Journal of the Acoustical Society of America*, *32*, 1046–1060. doi:10.1121/1.1908276
- Zwislocki, J. J. (1969). Temporal summation of loudness: An analysis. *The Journal of the Acoustical Society of America*, *46*(2B), 431–441. doi:10.1121/1.1911708

Received February 26, 2013

Revision received October 3, 2013

Accepted October 10, 2013 ■