AMPLITUDE ENVELOPE AND AUDIO-VISUAL PERCEPTION

THE ROLE OF AMPLITUDE ENVELOPE IN AUDIO-VISUAL PERCEPTION: TESTING THE EFFECT OF AMPLITUDE ENVELOPE IN SPATIAL VENTRILOQUISM

By DOMINIQUE BEAUREGARD CAZABON, B.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree Master of Science

McMaster University MASTER OF SCIENCE (2016)

Hamilton, Ontario (Psychology)

TITLE: The Role of Amplitude Envelope in Audio-Visual Perception: Testing the Effect of Amplitude Envelope in Spatial Ventriloquism

AUTHOR: Dominique Beauregard Cazabon, B.Sc. (McGill)

SUPERVISOR: Dr. Michael Schutz

NUMBER OF PAGES: x, 63

Abstract

The world is filled with richly diverse sounds which we are able to perceptually distinguish using a variety of properties. One of these properties is the amplitude envelope, or the intensity of a sound over time. While it is common in the real world for sounds to have time-varying amplitude envelopes, the majority of sounds used in perceptual research have time-invariant or unspecified amplitude envelopes. The aim of the present thesis is twofold. Because many of the studies using time-invariant or undefined envelopes make use of very short sounds (below 100 msec), the first experiment aimed to determine the duration required for discriminating among three different envelopes: flat (invariant), ramped (increasing in intensity over time), and damped (decreasing in intensity over time). In Experiment 1, participants took part in a 2-alternative forced choice, psychophysical staircase paradigm in which they indicated which of two envelopes they thought they were listening to. Results showed that, when telling ramped tones apart from either flat or damped tones, participants showed discrimination thresholds below 50 msec, while they had thresholds of approximately 75-80 msec when differentiating flat from damped tones. Because amplitude envelope has been shown to impact audiovisual integration and the perceptual system is sensitive to interaural envelope differences when localizing sounds, the second experiment aimed to determine whether amplitude envelope could modulate the visual bias present in spatial ventriloquism, an audiovisual illusion where the perceived location of a sound is influenced by the location of a visual stimulus. In Experiment 2, participants performed a psychophysical staircase task which measured their accuracy in localizing sounds with

flat and damped envelopes, with or without a simultaneous flash on the screen in front of them. Results showed that, at durations above the envelope discrimination thresholds found in Experiment 1 (83 msec), there was no visual bias on perceived location of the sound, while the bias was present at a duration below this threshold (16 msec). Together, these results add to the mounting evidence suggesting that amplitude envelope has profound and varied effects on our perception of sounds, and is an important property to consider when designing experiments.

Acknowledgements

I would first like to thank my supervisor, Dr. Michael Schutz, for providing me with the vast array of resources, both material and intellectual, necessary for the execution of this project. I would also like to thank my committee members, Dr. Sue Becker and Dr. Laurel Trainor, for their useful feedback at all stages of execution of this project. I would like to thank my labmates, Fiona Manning, Aimee Battcock, Anna Siminoski and in particular Lorraine Chuen for their support and helpful comments on my work.

This project would not have been possible without the help of the many wonderful undergraduate students I have had the chance to work with. Kimberly Germann's honours thesis is the basis for the experiment performed in Chapter 2 and her dedication to the project was nothing short of remarkable. I am also indebted to my undergraduate research assistants Marsha Natadiria, Erica Huynh and Brannon Senger for their dedication and their excellent work.

They say no man is an island, and the expression definitely goes for graduate students. I would not have been able to dedicate myself to this project without the unconditional support I received from my family: my mother, Louise Beauregard, my father, Robert Cazabon, and my sister, Anne-Marie Beauregard. I am also thankful for the support of two of my mentors from my time at McGill University, Evan Balaban and Stephen McAdams, and my friends, Liz Bellefleur-MacCaul, Mackenzie Churchill, Catherine Darwish, Kristin Langevin, Sarah McIlwaine, Kelyn Montano, Melissa

Rivosecchi, and Jeff Thiessen. Finally, a big thank you to the PNB ladies for inspiring me to keep going and lifting my spirits every step of the way.

This work was supported by NSERC (RGPIN/386603-2010), Early Researcher Award (ER10-07-195), Canadian Foundation for Innovation (CFI-LOF30101), and McMaster Arts Research board grants to Dr. Michael Schutz, and Ontario Graduate Scholarships and a FRQNT master's scholarship to me.

Table of Contents

List of Figures

List of Tables

Chapter 1

General Introduction

1) Navigating the auditory world

The auditory system performs complex signal-processing computations to allow us to navigate the world around us. Spatial hearing, the ability to locate sound sources, underlies the ability to determine the number of sound sources sources, orient to sound sources, and segregate auditory streams. Our ability for spatial hearing is also one of the factors enabling auditory phenomena such as the famous "cocktail party problem" introduced by Cherry (1953), which refers to our ability to keep up with one conversation at a cocktail party while ignoring the noise of other conversations. Unlike the visual and somatosensory systems, the location of auditory signals is not explicitly represented on the receptor surface; instead, sounds are mapped tonotopically along the basilar membrane, from the highest audible frequency at its apex to the lowest audible frequency at its base (Hudspeth, 2000). As such, the spatial representation of sounds is created through computations carried out subcortically and cortically.

The processing of auditory stimuli becomes even more sophisticated when the perceptual system must combine them with other modalities. We do not perceive sight, sound, smell, taste and touch in parallel; rather, the perceptual system combines them into a meaningful, unified experience. Combining these modalities is an efficient way of making sense of the large amount of information the perceptual system must handle. In fact, one of the ways the perceptual system binds stimuli from different modalities is by

combining stimuli which demonstrate cross-modal correspondences (Spence, 2011), in effect combining partially redundant information.

In the real world, the spatial and temporal information provided by our senses is correlated. In such conditions, cross-modal binding results in a functionally relevant representation of our surroundings. Studies show that the perceptual system combines the modalities in a way that would make our perception of multisensory percepts more accurate than our perception of single modalities (e.g. Alais & Burr, 2004). However, multisensory integration can also give rise to illusory percepts. For example, the perceived location of an auditory stimulus can be influenced by the location of a simultaneous visual stimulus. In an experimental setting, we have the ability to place information from the two modalities in conflict to give rise to these illusions. This process can reveal the obligatory mechanisms used by the perceptual system to integrate sight and sound. For example, when participants hear a voice uttering a syllable (e.g. 'ba') while viewing a video of lips uttering a different syllable (e.g. 'ga'), they reported hearing a syllable different from either of the two syllables they were perceiving (in this case, 'da') (McGurk & MacDonald, 1976). In spatial ventriloquism, a spatial conflict is introduced between the auditory and visual stimuli, resulting in participants demonstrating a visual bias in their perceived location of the sound source (e.g. Bertelson & Aschersleben, 1998).

2) Amplitude envelope

The pattern across neural firings on the basilar membrane act as a spectrogram. As they fire, they parse out what frequencies are present in the auditory stimulus over time, as well as their intensity. The frequency spectrum and intensity of the sound

combine to make up a wide array of attributes of the sound. One of these attributes is the amplitude envelope, which corresponds to the intensity of a sound over time (Vallet, Shore, & Schutz, 2014).

A common amplitude envelope we encounter in our everyday lives is the damped, or percussive, envelope. Damped envelopes have a sharp linear increase after onset, followed by an exponential decay. Impact events, such as items hitting the ground or striking notes on percussive music instruments, produce sounds with damped envelopes. Other naturalistic sounds, such as water flowing or the wind blowing, have amplitude envelopes that are hard to reproduce when generating artificial stimuli in the context of the laboratory. The amplitude envelopes used in perceptual research tend not to correspond to naturalistic sounds. After an experiment testing the envelope discrimination abilities of participants for three different amplitude envelopes (ramped, damped and flat), Germann (2016) asked participants to describe the different envelopes they heard. Damped tones were commonly compared to a xylophone. Flat envelopes, which have a sharp onset and offset, were often described by participants as sounding like "computer beeps" or "robot sounds," while ramped tones, which rise exponentially from onset and have a sharp decay, were rarely described as resembling sounds from the real world (Germann, 2016).

The study of auditory perception can be carried out using two main approaches. The first lies in determining the link between stimulus and sensation separate from context. The second, auditory scene analysis, is about determining how audition can provide us with a meaningful perception of the world around us. In the following

sections, I will outline how the two approaches have contributed to our understanding of amplitude envelope.

3) Audition in a vacuum: perception without context

Early studies of audition aimed to determine the structure and function of the auditory system by determining detection thresholds and fitting signal processing models to perceptual data. Such experiments have addressed questions about the effect of amplitude envelope on beat detection (Viemeister, 1970), amplitude modulation (AM) detection (Sheft & Yost, 1990; Viemeister, 1979), AM rate detection (Lee, 1994), and detection of changes in AM depth (Lee & Bacon, 1997). In this respect, this perspective has allowed us to determine our sensitivity to changes in different properties of amplitude modulation. In terms of perceptual properties, such studies have indicated that the amplitude envelope of a sound contributes to our perception of its timbre (Iverson & Krumhansl, 1993; Krimphoff, McAdams, & Winsberg, 1994), its duration (DiGiovanni & Schlauch, 2007; Grassi & Darwin, 2006; Schlauch, Ries, & DiGiovanni, 2001), its loudness (Neuhoff, 1998, 2001; Ries, Schlauch, & DiGiovanni, 2008; Stecker & Hafter, 2000), and, at very short durations, its pitch (Hartmann, 1978).

The main appeal of this approach is that it allows for good control over the experimental conditions. Methods such as psychophysical staircases and temporal order judgement tasks allow experimenters to measure perception without the possibility of factors like response bias explaining the resulting data. While they do not have to be, the auditory stimuli used in experiments are often synthesized, in which case experimenters can manipulate variables of interest while keeping components that are

not of interest constant across conditions. Such a degree of internal validity is desirable in many respects; however, high internal validity can come at the expense of external and ecological validity.

4) Auditory scene analysis: perception within context

While it is possible to listen for the properties of auditory stimuli (musical listening), we often listen to sounds in terms of what they tell us about the world around us (everyday listening) (Gaver, 1993). The human perceptual system does not operate within a vacuum devoid of context. Arguably, we perceive differences in loudness and amplitude modulation not just because they are important in and of themselves, but because these characteristics of sound tell us about the events producing them. As such, tones with amplitudes that ramp up or damp down over extended periods of time are perceived as looming and receding from the listener, and not just as getting louder or quieter. We even perceive looming tones as being closer to us than receding tones which end at the same loudness (Neuhoff, 1998), evidence of adaptive mechanisms that do not exist to represent sounds accurately, but that afford us better chances of survival from potential predators, for example (Neuhoff, 2001).

Auditory scene analysis focuses on defining perception. The purpose of scientists studying auditory scene analysis is to determine how the perceptual system "take[s] the sensory input and […] derive[s] a useful representation of reality from it." (Bregman, 1994, p.3). The perceptual system not only analyses sound in terms of auditory frequency over time, but also somehow creates a mental image of the world around us (Gaver, 1993).

Auditory scene analysis studies have shown that the amplitude envelope of a sound enables us to derive meaningful information about the event giving rise to a sound. Amplitude envelope is a reliable cue to determine whether a glass bottle broke or bounced off the ground (Warren & Verbrugge, 1984), the hardness of objects hitting a surface (Freed, 1990; Klatzky, Pai, & Krotkov, 2000), and the hardness of an object being struck (Giordano, Rocchesso, & McAdams, 2010).

Where psychophysical methods provides a high level of internal validity, studies investigating from an auditory scene analysis perspective give us insight into real-world perception, while still offering high internal validity. Auditory scene analysis experiments often use psychophysical methods to do so, while also ensuring that the stimuli used are ecologically valid.

5) Surveying sounds in perceptual research

While evidence exists for how amplitude envelope influences the perceptual quality and ecological validity of sounds, the importance of this characteristic is often overlooked in perceptual research (Gillard & Schutz, 2013; Schutz & Vaisberg, 2014). Of the papers sampled in the journal *Music Perception*, 35% omitted information about the amplitude envelope of sounds. In *Attention, Perception and Psychophysics*, 93% of studies used time-invariant (flat) envelopes. This lack of specification, as well as the dearth of articles using time-variant envelopes, raises questions about the degree to which these experimental outcomes generalize to natural sounds with time-varying amplitudes. We do not consciously perceive different characteristics of sound in parallel, but as a unified experience of auditory objects. All sounds have an amplitude envelope that

shapes the way we perceive them – any assessment of perception of other characteristics is incomplete without taking this into account. This is especially a problem when we use envelopes in research that do not resemble envelopes produced by real-life auditory events.

Researchers may argue that amplitude envelope may not affect results in many of these studies, as the use of short stimulus durations is common. In the *Attention, Perception and Psychophysics* survey, more than a third of sounds (36%) were below 100msec in duration (Gillard & Schutz, 2013). In audiovisual integration studies using non-speech sounds, the durations may be even shorter: a short informal survey of 20 articles about ventriloquism revealed that the median duration of stimuli was 16.7 msec, the duration of a single screen refresh on a standard computer monitor (Appendix A). As of yet, it is unknown if the amplitude envelope of sounds can even be perceived at such short durations.

6) Thesis objectives

The main objective of this thesis is to add to the body of literature allowing scientists to determine whether or not the amplitude envelope of sounds is an important variable to take into account in audition research. Specifically, the material in this thesis contributes to the literature on audiovisual integration by testing the effect of amplitude envelope on spatial ventriloquism. To this end, the following two chapters contain the results of two psychophysical experiments.

In Chapter 2, I present the results of an experiment determining the absolute duration threshold for the discrimination of three envelope types: flat, ramped, and damped.

In audiovisual experiments where the stimuli are devoid of contextual cues, short stimuli are more commonly used as they bind more easily. As such, determining the shortest duration at which participants could hear the difference between amplitude envelopes was necessary to determine the duration of the stimuli to be used in the present thesis. Results indicated that participants were able to discriminate envelopes at durations on the order of 60 msec on average. Participants performed better when discriminating between ramped tones and flat or damped tones, and performed worst when discriminating between flat and damped tones.

Using stimuli durations found in the previous study, chapter 3 details an experiment exploring the effect of amplitude envelope on the visual bias observed in spatial ventriloquism. Although the data failed to show visual bias for either of the envelopes used (flat and damped), the results shed insight on the narrow range of parameters used in psychophysical investigations of spatial ventriloquism.

In Chapter 4, based on the results of the two experiments and the literature review in this chapter, I make suggestions for how future experiments on amplitude envelope could be designed, and what knowledge can be gained from such investigations.

Chapter 2

Envelope discrimination

1) Background

The study of amplitude envelope discrimination can reveal many things about the auditory system. Individuals seem to instinctively associate certain envelopes with real-life objects. For example, participants are significantly more likely associate damped tones with events form the real world than flat or ramped tones (Germann, 2016), perhaps because damped or percussive tones are readily encountered in everyday life under the form of impact sounds. Flat tones are mostly encountered in contexts where sounds are synthesized, such as when hearing dial tones or computer sounds. Finally, ramped tones are rarely encountered in everyday life. While finding associations between envelopes and naturalistic sounds provides listeners with insight as to the perceptual object arising from these envelopes, the tendency to associate envelopes with naturalistic sounds does not result in better performance at discriminating between envelopes (Germann, 2016): participants who described sounds by comparing them to real-world events did not perform better than those who used descriptors relating to stimulus characteristics.

Greater understanding of envelope processing could also provide insight into auditory deficits in atypical populations. Individuals with dyslexia show deficits in detecting differences in the rate of change in amplitude envelope onsets, or rise time (Pasquini, Corriveau, & Goswami, 2007), beat perception (Muneaux, Ziegler, Truc, Thomson, & Goswami, 2004), and phonological representation (Goswami, Gerson, & Astruc, 2010; Richardson, Thomson, Scott, & Goswami, 2004). Such deficits are thought

to underlie the reading difficulties associated with dyslexia, showing that the processing of envelope contributes to such complex cognitive tasks as reading. Quantifying these deficits could lead to the creation of accessible diagnostic tests for early detection of dyslexia. Furthermore, understanding the nature of these deficits could shed insight both on normal mechanisms of envelope discrimination and the neurological abnormalities found in dyslexia.

From a researcher's vantage point, there is also much to gain from studying the discrimination of amplitude envelopes. Although an experimenter might hypothesize that amplitude envelope affects perception, this is impossible to examine this from the current literature as the majority of literature in acoustic perception fail to specify the envelopes used or focus heavily on time-invariant envelopes. Furthermore, should there be an effect of envelope, it would be difficult to determine the shortest duration at which amplitude envelopes should matter, because the duration below which we cannot discern one envelope from another remains unknown.

There exist a variety of theories that could explain amplitude envelope processing. As the duration of a stimulus increases, detection and discrimination thresholds decrease (Viemeister & Wakefield, 1991). As such, models of discrimination and detection are all based on integration of some sort. Leaky integration models posit that the envelope modulator is smoothed by the leaky integrator, after which a decision statistic is used to detect it. This decision statistic can be the standard deviation of the envelope (Viemeister, 1979), the ratio of the maximum of the envelope to its minimum (Forrest & Green, 1987), or the magnitude spectrum of the envelope (envelope spectrum model;

Akeroyd & Patterson, 1997). However, two different envelope modulators can have the same magnitude spectrum if they are temporally asymmetrical, time-reversed versions of one another. Sufficient evidence exists to show that listeners can discriminate between temporally asymmetrical modulators with different directions (Akeroyd & Patterson, 1997; Irino & Patterson, 1996). Thus, while the envelope spectrum model may describe the detection of asymmetrical amplitude modulation, it does not accurately account for the ability of listeners to discriminate between different directions.

Akeroyd and Patterson (1997) compared the ability of listeners to detect and discriminate temporal asymmetry in amplitude modulation. An amplitude envelope that has temporal symmetry is the same when played forward as when it is time-reversed; an amplitude envelope that is temporally asymmetrical is different when played forward than when it is time-reversed. In their experiments, they used sinusoidal amplitude modulation (SAM) as well as downward-sloping sinusoidal amplitude modulation (D-SAM) and upward-sloping sinusoidal amplitude modulation (U-SAM), the latter two being temporally asymmetrical and mirroring one another, and the former having temporal symmetry. They found that the detection of amplitude modulation did not depend on the temporal direction of the modulation; as such, the envelope spectrum model described amplitude modulation detection well. On the other hand, the discrimination of amplitude modulation slope direction did depend on the direction of modulation and was not well described by the envelope spectrum model. Additionally, they noted that all their results, both for detection and discrimination, were accurately represented by a model using temporal processes.

The aim of this experiment is to determine the shortest duration at which listeners can discriminate between flat, damped, and ramped envelopes, as well as whether certain envelopes are harder to tell apart from others. The duration threshold is difficult to predict, because most studies use many repetitions of SAM tones to estimate their thresholds, while the present study focuses on discrete envelopes. However, several predictions can be made as to possible differences in thresholds depending on the envelopes being discriminated. Discrimination of U-SAM noise from D-SAM noise is better than that of either noise from SAM noise (Akeroyd & Patterson, 1997), which suggests that ramped and damped tones may be the easiest to tell from one another. Additionally, for pure tones, listeners are better able to detect differences in linear onsets than offsets (Van Heuven & Van Den Broecke, 1979). While the ramped tones used in this experiment rise exponentially from the onset, which were not compared to linear onset, because flat tones and damped tones both have similar linear onsets, they may be more difficult to tell apart than either of those tones when compared to ramped tones. By the same token, the difference in offsets for flat and damped tones may not be as important in differentiating them.

2) Methods

a) Subjects

Thirty-eight participants (32 female, 6 male; mean age 19) from the McMaster University Psychology Participant Pool took part in one 1-hour session. As compensation, participants received $10 or one course credit in introductory psychology. The data of 5 participants were discarded because they were unable to complete the experiment within the hour session, leaving the data of 33 participants (28 female, 5 male; mean age 20).

All participants filled a consent form and were screened for normal hearing thresholds prior to participating in the experiment. This research was approved by the McMaster University Research Ethics Board.

b) Auditory stimuli

The experiment made use of flat, ramped and damped tones with durations ranging from 15 msec to 500 msec in 5 msec increments generated using MATLAB 7 (The MathWorks Inc., Natick, Massachussetts). All tones had a pure tone sinewave carrier with a frequency of 2 kHz. This high frequency was used because higher pure tone frequencies contain more cycles than lower frequencies for the same duration, providing more information about the amplitude envelope.

To ensure that stimuli of the same duration had the same loudness, flat tones had 5 msec linear ramp onset and offset, and a sustained amplitude of half the maximum amplitude of the other tones (Figure 1). The damped tones had a linear ramp onset of 5 msec and an exponential decay described by the equation $x(t) = e^{(-5t/T)}$ where t is the time and T is the duration (Schlauch, Ries, & DiGiovanni, 2001). The ramped tones were time-reversed equivalents of the damped tones. The intensities of the stimuli (A-weighted, fast impulse) were 78.5 dB SPL for the flat tones, 78.0 dB SPL for the damped tones, and 79.0 dB SPL for the ramped tones.

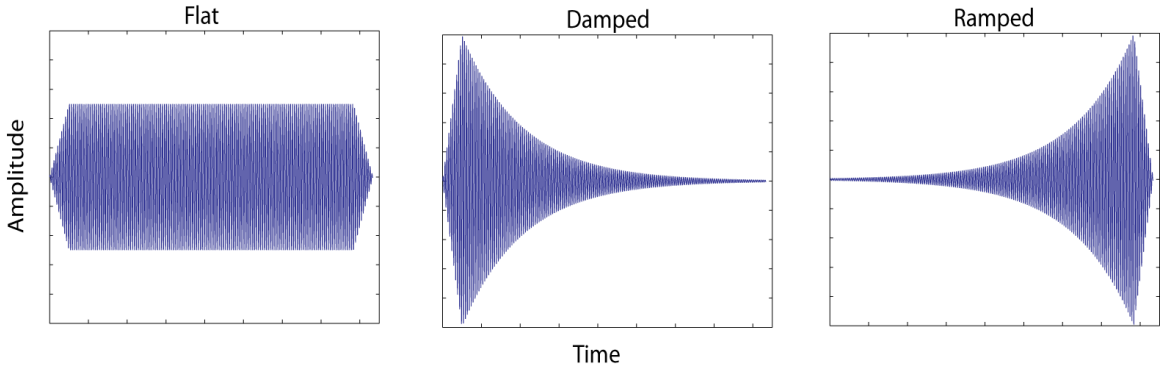

Figure 1: Amplitude envelopes of the stimuli.

c) Experimental set-up

The experiment took place in a sound-attenuating booth (Industrial Acoustics Company Inc., Bronx, New York). Experimenters tested participants' hearing thresholds using a MAICO MA-25 audiometer. Each subject sat at a desk in front of a 19" Dell computer monitor. They listened to auditory stimuli through Sennheiser HD-280 Pro 64 Ω headphones connected to a Scarlett 2i2 audio interface. At each trial, participants entered their responses using a Mac keyboard. An iMac computer (2.7GHz, Intel Core i5, 8GB 1600 MHz DDR3) controlled presentation of stimuli, display of instructions on the computer monitor and recording of responses through a script programmed in Python using PsychoPy2 (v1.80.03) (Peirce, 2007).

d) Design

Participants performed a 2-alternative forced-choice task (2AFC) in which they had to indicate which of 2 possible envelope categories a presented sound belonged to. Stimuli were presented in two interleaved 1-up 3-down psychophysical staircases, providing discrimination thresholds where participants can correctly determine the envelope 70% of the time (Levitt, 1971). Each staircase was made up of auditory stimuli

with a flat, ramped or damped amplitude envelope. In each trial, one of the two staircases was selected at random, and participants heard a stimulus from this staircase. For each staircase, the first stimulus presented had a duration of 410 msec. If the participant correctly identified the category of the sound three times, the next duration for that staircase was decreased by one step. If the participant provided a single incorrect response, the next duration for that staircase was increased by one step. The step size changed according to the number of reversals already performed by the staircase: the staircase increased and decreased by 40 msec before the first reversal, 20 msec before the second, and 5 msec for the remaining four reversals. Each pair of interleaved staircases ended once six reversals were recorded per staircase.

e) Procedure

Participants signed a consent form, then filled a short survey containing questions about their age, gender, history of hearing problems, and languages spoken, as well as the Ollen Musical Sophistication Index (OMSI) questionnaire. Participants entered the sound-attenuating booth. The experimenter performed a hearing test to ensure their hearing was within 20dB of the required threshold of normal hearing for each of the frequencies tested (ISO, 1998; Martin & Champlin, 2000).

Experimenters then explained the procedure to participants while instructions were displayed on the monitor. Participants were told they would familiarize themselves with three different categories of sounds, after which their ability to classify sounds into these categories would be tested. They were warned that catch trials, in which a sound different from the three categories would be played (alternating E6 (1319 Hz) and C6

(1047 Hz) notes), would occur at random. They were instructed to respond to catch trials by pressing the spacebar.

In the initial familiarization phase, participants could listen to the three different categories of tones as many times as they needed, choosing which sounds they wanted to hear by pressing buttons on the keyboard: 1 for flat tones (Category 1), 2 for damped tones (Category 2), and 3 for ramped tones (Category 3). For each trial in the testing phase, a sound from one of the three categories was played 3 times, after which participants were required to indicate which category they thought it belonged to by pressing keys 1, 2, or 3. Participants were able to complete a short "warm-up" block of the testing phase with the experimenter in the room to make sure the instructions and controls were clear. After this, the instructor left the sound booth and the participants were asked to complete the remainder of the experiment on their own.

The experiment was split into 3 different blocks, in which participants were asked to compare only 2 of the 3 different categories of sound at a time: Flat vs Ramped, Damped vs Ramped, and Flat vs Damped. Because the experiment tested separately for 6 comparisons, it can be seen as a "standard" versus "comparator" tone task where the standard is the first tone in a given block. Each block was composed of a familiarization screen allowing the participants to listen to the two categories being tested once again, followed by three sets of interleaved staircases. Participants had the possibility to take breaks between the staircases and blocks and advanced through the different parts of the experiment using the spacebar. Block order was counterbalanced across participants using complete counterbalancing.

Once the 3 blocks were completed, participants completed a short post-experiment survey where they answered questions about their perception of the sounds they heard, after which the experiment was complete.

3) Results

Data were processed and analysed using R version 3.2.3. Discrimination thresholds were calculated by obtaining the mean of the last four reversals for each staircase. Table 1 shows descriptive statistics for each envelope and the envelope of the sounds they were compared to.

Envelope	**Flat**		**Damped**		**Ramped**	
Comparison Envelope	Damped	Ramped	Flat	Ramped	Flat	Damped
Mean (msec)	74.3	46.8	77.5	34.8	42.3	43.1
SD (msec)	81.0	52.8	84.4	38.5	41.4	47.7
SEM (msec)	14.1	9.2	14.7	6.7	7.2	8.3
Median (msec)	25.5	23.1	25.0	21.9	24.4	21.9
Skewness	1.2	2.3	1.1	3.6	2.6	2.4

Table 1: Descriptive statistics by envelope and comparison envelope.

a) Effect of envelope

Bootstrap test statistics comparing the thresholds for comparison envelopes at each level of the envelope factor were calculated. To account for multiple comparisons, a Bonferroni-corrected alpha level of 0.017 (0.05÷3) was used. When the tones were flat, participants were able to discriminate shorter tones better from ramped tones than from damped tones ($p=.0025$). For damped tones, participants discriminated shorter tones from

ramped tones better than from flat tones (p<.001). Finally, for ramped tones, there was no difference in thresholds when discriminating against flat or damped tones (p=.5425).

Non-parametric bias-corrected and accelerated bootstrap confidence intervals (95%) were computed using the bcanon function part of the bootstrap package in R (Efron & Tibshirani, 1994). As shown in Figure 1, participants had lower discrimination thresholds when they differentiated ramped tones from other tones.

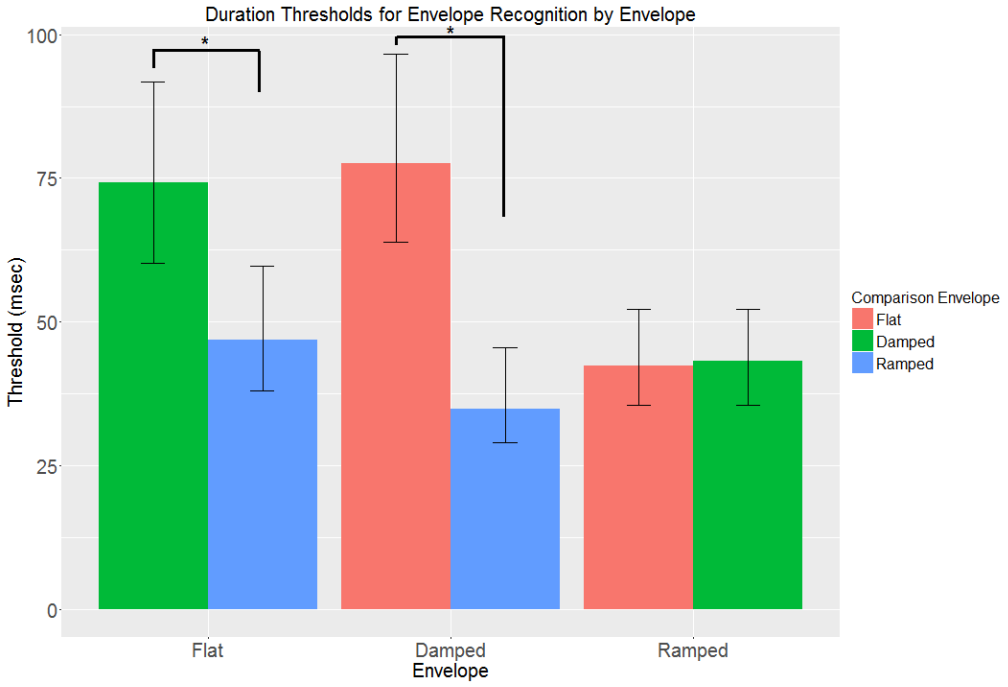

Figure 2: Means and bootstrap 95% confidence intervals for discrimination thresholds. Asterisk (*) denotes p < 0.05.

b) Correlation analyses

Overall mean thresholds were calculated for each participant. These thresholds were not significantly correlated with OMSI scores, Pearson's R(33)=-.020, p=0.92. The

number of languages spoken by the participants and the overall duration thresholds were not significantly correlated, Pearson's $R(33)=0.154$, $p=0.39$.

4) Discussion

Participants obtained lower thresholds when discriminating tones with exponential rising onsets (i.e. ramped) from those with sudden onsets (flat, damped). However, they did not exhibit better performance when discriminating tones with exponentially decaying offsets (i.e. damped) from those with sudden offsets (ramped, flat). This is consistent with the existing literature; for pure tones, detection of changes in linear onsets is better than for offsets (Van Heuven & Van Den Broecke, 1979). While the present study featured sounds with both linear and exponential ramping and damping portions, it is apparent that the exponential onset of the ramped tone is markedly different from the linear onsets of the damped and flat tones. Because both damped and flat tones have a linear ramp-up portion from onset, they may have been more difficult to tell apart. Because ramped tones are not encountered in the real world, these results suggest an association – it was easy to discriminate non-real world sounds from real-world sounds.

These results differ from the envelope discrimination data of Akeroyd and Patterson (1997). Using temporally asymmetrical sinusoidal amplitude modulated tones, the authors found that listeners could discriminate best between the upward-sloping SAM tones and the downward-sloping SAM tones, whereas they did not perform as well when discriminating between symmetrical SAM tones and either of the asymmetrical SAM tones. These results suggest that the more "different" envelopes are in terms of temporal

symmetry, the easier they are to discriminate. In the present experiment, the ramped and damped tones would be the most different, with the flat tone being equally different from either of the former two. In this respect, we would expect participants to perform best when discriminating ramped and damped tones. However, the current data do not follow this trend: listeners were better whenever they had to discriminate a ramped tone from another, regardless of the envelope of the other tone. Many factors may play into this. Participants in the present experiment had to discriminate between envelopes that did not repeat periodically, as is usually the case with SAM tones. Rather, the stimuli were repeated three times, which may not have been enough to provide participants with a sense of the difference between flat and damped tones. It seems that, especially at shorter durations, the onset of the tones may be the main cue to determine the type of envelope participants are listening to. Additionally, Akeroyd and Patterson (1997) used noise carriers, which may provide more information about the amplitude envelope of stimuli than the 2 kHz sine wave carrier used here. With sine wave carriers, the information about amplitude modulation can only be derived from the maxima and minima of the sine wave, whereas noise carriers contain such maxima and minima for an array of frequencies. If the difference in performance is indeed the result of this difference, we would expect performance for pure tones to increase with the frequency of the carrier, as a higher frequency would result in more instances of cycling through the maximum and minimum within the stimuli. Similarly, we would expect performance for noise carriers to be better for broadband noise than narrow-band noise, and high pass-filtered noise to be better than low pass-filtered noise.

These results add to the literature on the perception of temporally asymmetrical amplitude envelopes. In addition, they provide a duration threshold for the perception of differences in discrete amplitude envelopes, which is useful in the design of other perceptual experiments. Future directions include the study of such discrete envelopes with carriers containing more frequencies, such as white noise, to mimic the spectra of sounds encountered in the real world.

Chapter 3

Audio-visual integration

1) Background

Humans have the ability to combine information from the 5 senses into a single meaningful percept. Doing so requires relating information across different modalities. As a consequence of integration, one's perception of a stimulus in one modality can often be changed by what is experienced in another modality. Several striking examples of this exist in instances of audiovisual integration. Audiovisual speech, for example, can improve speech comprehension in a noisy environment (Rudmann, McCarley, & Kramer, 2003). Other classic examples include the McGurk effect (McGurk & MacDonald, 1976), the ventriloquism effect (e.g. Bertelson & Aschersleben, 1998) and the Colavita effect (Colavita, 1974). These illusions showcase the trade-offs between audition and vision that result from integration of the two modalities. Describing and understanding these trade-offs is an integral part of understanding how the perceptual system combines information from the two modalities. Further, qualifying these trade-offs under different audiovisual conditions can help us understand if the perceptual system integrates auditory and visual information optimally (e.g. Alais & Burr, 2004; Sato, Toyoizumi, & Aihara, 2007), a widely accepted theory of audiovisual integration.

a) Trade-offs in audiovisual integration

Basic studies of audiovisual integration aim to characterise trade-offs in two dimensions: the spatial dimension (where did something happen?) and the temporal

dimension (when did something happen?). The study of the spatial dimension mainly revolves around an illusion dubbed the ventriloquism effect, which results in the perception of spatial congruence when, in reality, an auditory and visual stimulus occurring simultaneously are in different locations. In this illusion, vision is weighted more strongly than hearing when localizing events, biasing participants' location judgements of sound sources towards the location of the visual event.

The study of the ventriloquism effect began with behavioural paradigms, such as Jackson's (1953) study featuring an array of steaming kettles and whistles. Participants in this experiment were asked to determine where the whistling sounds came from. They tended to localize the sound of the whistle as coming from a steaming kettle when the two were closer together, with this effect decreasing with larger separations between the auditory and visual stimuli. The authors described this phenomenon as a consequence of "visual dominance," concluding that vision predominates in localization tasks. Such paradigms, while common in the beginnings of studies on the ventriloquism effect, were flawed, as participants' responses were likely to be affected by response bias (Choe, Welch, Gilford, & Juola, 1975). Auditory localization is more difficult than visual, so participants are likely to point to visual stimuli because they are easier to localize, dismissing their auditory perception in favour of pointing towards an easier event to locate. Alternatively, participants may demonstrate response learning. If a participant is asked to point to an auditory source before and after the introduction of a conflicting spatial stimulus, it is possible that participants' responses may simply be biased towards the visual stimulus after introduction of the stimulus, especially if that is what they sense

the researchers are hoping to observe. Fortunately, psychopysical methods can address the issue of response bias, either through statistical means (Choe et al., 1975) or through the use of non-transparent tasks (Bertelson & Aschersleben, 1998).

Studies have begun to uncover the neurological mechanisms underlying the ventriloquism effect. A functional magnetic resonance (fMRI) study by Bonath et al. (2007) revealed that the planum temporale (PT) in the auditory cortex is activated the same way by both real sound shifts and illusory shifts caused by the ventriloquism effect. In addition, electroencephalography (EEG) studies show these same illusory sound shifts can elicit the mismatch negativity (MMN), indicating that the change in perception happens early on in perception (Colin, Radeau, Soquet, Dachy, & Deltenre, 2002; Stekelenburg, Vroomen, & de Gelder, 2004). Together, these results suggest that the shift in perception occurs before the level of the cortex, likely in subcortical structures such as the superior colliculus (Alais, Newell, & Mamassian, 2010).

In the temporal dimension, audiovisual integration allows the perceptual system to perceive synchrony between auditory and visual stimuli, even when the two may have occurred at slightly different times. This allows the perceptual system to resolve time differences introduced by the stimulus being at farther distances from the perceiver, for example, since the speed of light is over 88 thousand times greater than the speed of sound. When viewing and hearing stimuli up close (less than one meter), time differences related to traveling time are negligible. At these distances, time differences are introduced by differences in transduction speeds between vision and hearing (King & Palmer, 1985). Evidence suggests that the brain integrates sound and sight so long as

they fall within a movable time window which is calibrated according to visual information about the distance at which a given stimulus has occurred (Sugita & Suzuki, 2003).

Other research on the perception of synchrony highlights the discovery of a temporal analogue to the ventriloquism effect. Dubbed "temporal ventriloquism," this phenomenon occurs when a visual event becomes perceptually aligned with the onset of an auditory event despite the two being presented asynchronously. In this case, audition dominates over vision, the opposite of what happens in spatial ventriloquism. Temporal ventriloquism can, for example, change the perceived time of arrival of visual events (Morein-Zamir, Soto-Faraco, & Kingstone, 2003), influence the perceived rate of flickering of a visual stimulus (Recanzone, 2003), or affect one's performance on a temporal order judgement (TOJ) task involving audiovisual speech stimuli (Vatakis & Spence, 2007).

b) Optimal integration

How can we explain the trade-offs we observe between audition and vision? Why does vision predominate in spatial tasks, and audition in temporal tasks? An early theory explaining these phenomena was the modality specificity hypothesis (Welch & Warren, 1980). The hypothesis states that the perceptual system will resolve discrepancies between two modalities by favouring the modality that offers more precision. As such, in the case of spatial discrepancies, vision offers very high spatial acuity, sometimes even less than one degree of visual angle (Cavonius & Robbins, 1973). On the other hand,

human acuity in acoustic space is worse than for vision (Recanzone, Makhamra, & Guard, 1998) and varies significantly depending on the characteristics of the auditory stimulus. The localization acuity of Old World monkeys on the azimuthal plane varied between 18 degrees for a bandwidth 250 Hz and 4 degrees for a bandwidth of 8000 Hz (Brown, Beecher, Moody, & Stebbins, 1980). The intensity of sounds can also affect localization acuity (Miller & Recanzone, 2009; Sabin, Macpherson, & Middlebrooks, 2005). When sounds have a low intensity (14 dB), humans have a localization threshold of approximately 5 to 10 degrees on the azimuthal plane, and 10 to 15 degrees on the elevational plane, while these thresholds fall below 5 degrees when the intensity is higher (30 dB) (Su & Recanzone, 2001). In the temporal dimension, audition provides better acuity than vision. Humans with normal hearing can detect differences in auditory signals as small as 5 milliseconds (Michalewski, Starr, Nguyen, Kong, & Zeng, 2005), but visual temporal acuity is on the order of 20 milliseconds (Kietzman, 1967). Hence, audition predominates over vision in resolving temporal discrepancies.

More recently, experimenters have begun explaining observed trade-offs between vision and audition in terms of Bayesian probabilities. According to this view, the brain combines different modalities optimally given the relative noisiness of each modality (e.g. Ernst & Banks, 2002). In other words, the perceptual system weighs the information provided by each modality according to the reliability of the information it provides about the distinction one is trying to make. Because visual information about location is generally more reliable than auditory information about the same characteristic, vision predominates in audiovisual localization tasks, and localization judgements are biased

towards visual targets in the case of spatial discrepancies. This hypothesis may seem very similar to the modality precision hypothesis, but the two diverge on a crucial point: in optimal integration, the weights attributed to each modality can change dynamically in different contexts. For example, when the visual signal is made less reliable through blurring, greater weight is attributed to information provided by the auditory stimulus (Alais & Burr, 2004). For example patient whose auditory acuity is greater than their visual acuity as a result of Balint's syndrome demonstrated a similar performance with no degradation of the visual stimulus (Phan, Schendel, Recanzone, & Robertson, 2000). Additionally, computational models using ideal observers yield results similar to human participants (Sato et al., 2007).

An issue with the optimal integration literature as it currently stands is its generality as there has been a lack of diversity in the stimuli and paradigms it has employed. Traditionally, ventriloquism was studied using a variety of experimental designs, from the simpler combinations of sound bursts and light flashes (e.g. Bertelson & Aschersleben, 1998; McGrath & Summerfield, 1985; Zampini, Shore, & Spence, 2003) to more ecologically valid stimuli such as audiovisual speech (e.g. Vatakis & Spence, 2007) and musical instruments (Schutz & Lipscomb, 2007). The ventriloquism effect has even been studied in dart-poison frogs (Narins, Grabul, Soma, Gaucher, & Hödl, 2005). In addition, researchers have investigated the effect of many characteristics on audiovisual integration, such as spatiotemporal correspondence, and semantic and synaesthetic congruency (Spence, 2007, 2011). Studies using Bayesian frameworks, on the other hand, have been limited to simple stimuli such as white disks appearing on computer

screens and pure-tone beeps (Shams, Ma, & Beierholm, 2005), or clicks and Gaussian blobs (Alais & Burr, 2004). Even given the simplicity of the cues utilized, there is a wide gulf between the sensory integration we observe in a lab setting and that which occurs in real life. For example, Alais & Burr (2004) simulated the lateralization of their auditory stimuli by introducing an interaural time difference, omitting the intensity differences and spectral changes one would also encounter if one were to hear the very same click in the real world. The authors admit, in a later paper, that using only interaural time differences reduces auditory spatial acuity significantly; by their estimate, to about one sixth of our acuity under normal hearing conditions (Burr & Alais, 2006). The authors expect that, given normal listening conditions, the visual stimulus would simply have to be less degraded before audition would start to predominate in location judgements. While this makes sense in theory given their findings, there is no way of knowing this without repeating the experiment with stimuli reflecting listening and viewing conditions closer to real life. In addition, studies often feature the use of computational models and ideal observers rather than human subjects (e.g. Sato et al., 2007). Ideal observers can provide the means of testing out new theories, as much insight can be gained from comparing their performance at a task to that of human participants. That being said, the resemblance of human performance to that of an ideal observer's does not confirm that both have achieved these results through the same process.

The constrained nature of the stimuli frequently used in optimal integration make it difficult to determine whether Bayesian models of audiovisual integration are capable of describing integration of more complex sets of stimuli. Other psychological studies

show that audiovisual integration changes significantly depending on factors such as visual perceptual grouping (Sanabria, Soto-Faraco, Chan, & Spence, 2004), but it is difficult to determine how a Bayesian model would account for such a factor. This lack of stimulus complexity is largely due to the fact that optimal integration research is still in its beginnings. Now that basic frameworks have yielded promising results, studying optimal integration using more complex stimuli can help answer basic questions about the degree to which this framework generalizes to natural situations. First, doing so would help to determine if optimal integration can explain more ecologically valid instances of audiovisual integration. Second, doing so would allow for comparison to experiments carried out to test non-Bayesian hypotheses, an issue certain researchers have with Bayesian hypotheses (Bowers & Davis, 2012). Increasing the complexity of the stimuli incrementally would allow for the gradual building of models describing audiovisual perception. Initially, this could be done by determining the effects of manipulating new auditory or visual characteristics using well-established paradigms. This would allow for the determination of how different aspects of audiovisual stimuli affect trade-offs between modalities – to obtain a greater understanding of how the perceptual system weighs visual and auditory cues in more complex situations. In terms of auditory characteristics, manipulating the amplitude envelope of sounds provides an interesting first step in determining what auditory cues are considered more or less informative by the perceptual system, and how it does so.

c) Inter-aural differences

The auditory system makes use of a variety of cues in order to localize sounds. These cues can be separated into monaural cues, which arise from analysing the signal coming into a single ear, and binaural cues, which arise from comparing the signal as it arrives to each ear. Monaural cues are useful to determine the elevation of sounds, while binaural cues are useful to determine the location of a sound source in the horizontal plane. Binaural cues can be split into interaural time differences (ITDs) and interaural level differences (ILDs). Which cues the brain utilizes to locate sounds depends on the range of frequencies of the sound; according to Rayleigh's "duplex theory" of spatial hearing (1907), ILDs are used for high frequency sounds (above 1,500 Hz), as these sounds cast a sufficient "head shadow" to yield a perceivable level difference between both ears. Low frequency sounds, on the other hand, do not cast such a sufficient head shadow, and therefore the perceptual system uses ITDs to locate them. Rayleigh proposed that ITDs could not be used to determine the location of high frequency sounds, because the oscillations are too fast to determine which ear is leading and which is lagging. However, researchers still found that participants were sensitive to interaural time differences for sounds with complex waveforms made up of high frequencies, such as sinusoidal amplitude modulated tones, transients, and bands of noise (Bernstein, 2001), albeit less sensitive to time differences for these sounds than for those of similar sounds with low frequencies. Further studies revealed that humans were sensitive to time differences of the amplitude envelope of the waveform, rather than the fine structure. Participants exhibited the same performance when there was a time difference for the

envelope, but not the waveform of sounds, as when the entire sound was delayed in one ear (Nuetzel & Hafter, 1976). The opposite is true for complex waveforms made up of low frequencies: participants are sensitive to differences in fine structure, but not envelope cues (Henning, 1980; Henning & Ashton, 1981).

Amplitude envelope has also been shown to play a role in audiovisual integration. Schutz & Lipscomb (2007) created an experiment in which participants had to watch a video of a marimbist striking a note and indicate how long they perceived that note to be. Independent of the acoustic duration of the note, participants' duration estimates depended on the gesture of the marimbist after striking the note. A longer gesture yielded longer duration judgements and a shorter gesture yielded shorter duration judgements. Further experimenting revealed this visual influence on an auditory duration assessment was specific to percussive sounds, and does not persist with sustained sounds (Schutz & Kubovy, 2009). In another study observing the effect of amplitude envelope on audiovisual integration, Grassi & Casco (2009) found that participants viewed a video differently depending on the sound accompanying it. The video, depicting two discs starting at opposite points and moving towards one another in a straight line, was accompanied by either an impact-similar sound, with an abrupt attack, or an impact-dissimilar sound, with a gradual attack. The impact-similar sound yielded a percept of the two circles bouncing against one another, while the impact-dissimilar sound yielded a percept of the circles running through each other. While these studies address the effect of amplitude envelope on audiovisual integration more in the lens of the unity assumption (could the visual give rise to this sound?) rather than optimal integration, they remain

nonetheless evidence that amplitude envelope plays a role in how audiovisual events are perceived. This, combined with the fact that people are sensitive to interaural time differences in the envelope of complex waveforms, opens the door for questions about how amplitude envelope could affect the weighing of audiovisual stimuli in optimal integration. Specifically, our sensitivity to envelope ITDs makes studying cases of audiovisual integration in the spatial domain, such as the ventriloquism effect, particularly interesting.

The following experiments aim to determine whether amplitude envelope can modulate the magnitude of visual bias in spatial ventriloquism by comparing this visual bias for two amplitude envelopes: flat and damped. If audiovisual integration is indeed optimal, we expect to see differences in visual bias only if one of the two envelopes results in reduced noise and therefore better localization. While the stimuli used have the same duration, because damped tones sound much shorter than flat tones of the same duration, it is expected that the flat tones would be localized better. If such a difference in localization thresholds is found in conditions without any visual input, then participants should show less visual bias for the better-localized sound. The psychophysical method used allows for measuring the ability of participants to localize sounds while eliminating the possibility of response bias. Because the participants are localizing sounds from the left and the right, and the visual event takes place on the midline, spatial ventriloquism would result in participants perceiving the sound source as closer to the midline than it really is. As such, ventriloquism would result in participants having thresholds that are further from the midline for the audiovisual condition than for audio alone.

2) General methods

a) Experimental set-up

The experiment took place in a sound-attenuating booth (Industrial Acoustics Company Inc., Bronx, New York). First, experimenters tested participants' hearing thresholds using a MAICO MA-25 audiometer. Then, each subject sat at a desk in front of a 19" Dell computer monitor. Participants heard the auditory stimuli through two Focal Alpha 65 studio monitors placed 30 cm on either side of the computer monitor at ear level. The monitors were connected to an iMac computer (2.7GHz, Intel Core i5, 8GB 1600 MHz DDR3) by a Scarlett 2i2 audio interface. A microfiber curtain hung from the ceiling of the sound booth hid the speakers from view of the participants, with a square cut out to allow the computer screen to remain visible. Participants placed their head on a chinrest to ensure consistent head position. On each trial, they entered their responses using a Mac keyboard. A script programmed in Python using PsychoPy2 (v1.80.03) (Peirce, 2007) controlled the presentation of stimuli, display of instructions on the computer monitor, and recording of responses.

b) Auditory stimuli

Auditory stimuli were generated using a 60 kHz sampling rate in MATLAB 7 (The MathWorks Inc., Natick, Massachussetts) and spatialized along the azimuth by introducing onset phase differences between the left and right channels of the wav files. The differences ranged from 40 µsec to 960 µsec, increasing by 40 µsec from 40 µsec to 400 µsec and by 80 µsec from 480 t0 960 µsec. Experiment 1 used sinewave bursts, while experiments 2.1 and 2.2 used flat and damped tones.

c) Design

The experimental design was based on that of another spatial ventriloquism study by Bertelson & Aschersleben (1998). Participants performed a 2-alternative forced-choice task (2AFC) in which they had to indicate whether sounds came from the right or left of the midline of the computer monitor. Because the visual bias in spatial ventriloquism can be small, stimuli were presented in two interleaved 1-up 1-down psychophysical staircases, one starting from the left side, and one starting from the right. In each trial, one of the two staircases was selected at random, and the participant heard a stimulus from this staircase. For each staircase, the first stimulus presented had a phase difference of 640 μsec. If the participant correctly identified the category of the sound three times, the phase difference of the next auditory stimulus for that staircase was decreased. If the participant provided a single incorrect response, the next phase difference for that staircase was increased. The step size changed according to the number of reversals already performed by the staircase: the staircase increased and decreased by 4 positions before the first reversal, 2 positions before the second, and 1 position for the remaining four reversals. Each pair of interleaved staircases ended once eight reversals were recorded per staircase.

d) Procedure

At the beginning of the first session, participants signed a consent form, then filled a short survey containing questions about their age, gender, history of hearing problems, languages spoken, as well as musicianship. Participants entered the sound-attenuating booth in which they were screened to ensure their hearing was within 20dB of the

required threshold of normal hearing for each of the frequencies tested (ISO, 1998; Martin & Champlin, 2000).

Experimenters then explained the procedure to participants while instructions were displayed on the monitor. Participants were told that catch trials, in which a red dot would flash on the screen, would occur at random, and to respond to catch trials by pressing the space bar. For each trial, a sound from one of either the left or right staircase played 3 times, with presentations separated by an interval of 800 msec. Participants completed a short "warm-up" block of the testing phase with the experimenter in the room to make sure the instructions and controls were clear. After this, the instructor left the sound booth and the participants completed the remainder of the experiment on their own. At the end of the final session, participants filled a short survey on how difficult they found the experiment, whether they fell asleep, and if they had any strategies for localizing the sounds.

Experiment 2.0

1) Methods
    a) Subjects

Twenty-one participants (18 female, 3 male, mean age 20) from the McMaster University Psychology Participant Pool took part in two 1-hour sessions. The data of four participants were omitted from analysis, as they failed to complete the sessions within the hour slot, yielding incomplete data sets. As such, the data analysed is comprised of the data from seventeen participants (14 female, 3 male, mean age 19). Participants received

$10 per session for a total of $20. All participants signed a consent form and were screened for normal hearing thresholds prior to participating in the experiment. This research was approved by the McMaster University Research Ethics Board.

b) Auditory stimuli

The experiment made use of 16 msec pure tone pulses with a 2 kHz sinewave. Tones had an amplitude envelope described by the equation $x(t) = sin(\pi t/T)$, where t is the time and T is the duration. The pulses had an A-weighted, fast impulse intensity of 67.2 dB SPL.

c) Conditions

The script manipulated the independent variable of audiovisual condition (synchronous flash, SF versus no flash, NF). For the synchronous flash condition, a white dot with a 1.5 cm diameter flashed on the center of the screen simultaneously with the presentation of each auditory stimulus. For the no flash condition, a black screen accompanied the presentation of the stimuli. Each 1-hour session was split into 2 different blocks, one for each condition. Each block was composed of three successive pairs of interleaved staircases. Participants had the possibility to take breaks between the pairs of staircases and blocks and advanced through the different parts of the experiment using the spacebar. Block order was counterbalanced across sessions such that participants perform each block order once. Order of the sessions was counterbalanced across participants.

2) Results and Discussion

Data were processed and analysed using R version 3.2.3. Discrimination thresholds were calculated by obtaining the mean of the last six reversals for each staircase. Table 2 shows descriptive statistics for each condition.

Table 2: Descriptive statistics for localization thresholds by audiovisual condition.

Condition	**No Flash (NF)**	**Synchronous Flash (SF)**
Mean (degrees)	17.3	19.8
SD (degrees)	4.37	5.03
SEM (degrees)	1.06	1.22
Median (degrees)	5.4	11.8
Skewness	1.09	0.569

The bootstrap test statistic comparing the thresholds for each condition was significant ($p = 0.003$). Participants were better able to determine the location of sounds that were closer to the center in the NF condition than the SF condition. Non-parametric bias-corrected and accelerated bootstrap confidence intervals (95%) were computed using the bcanon function part of the bootstrap package in R (Efron & Tibshirani, 1994). As shown in Figure 3, participants demonstrate a visual bias of the perceived auditory location in condition SF, resulting in thresholds for location discrimination being further from the midline.

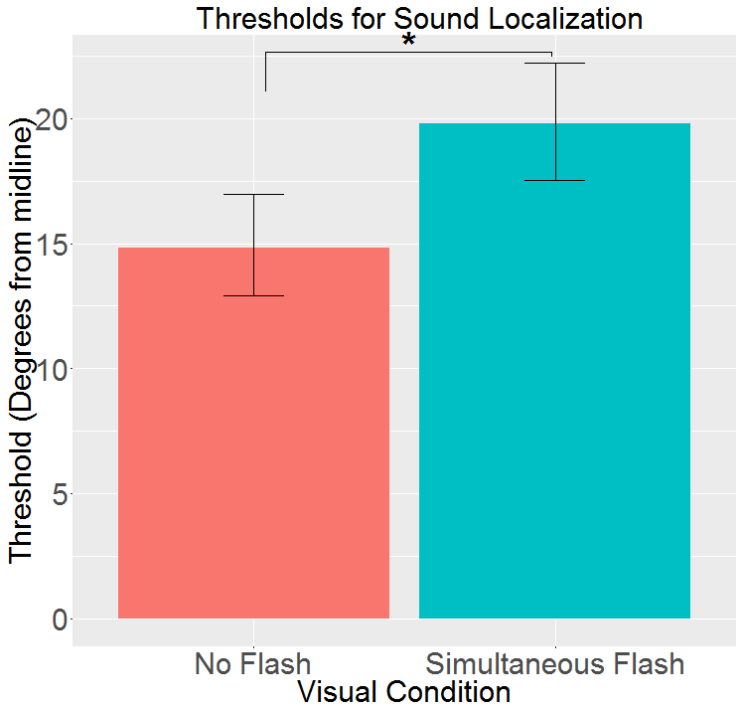

Figure 3: Location discrimination thresholds by audiovisual condition.

Error bars denote bootstrap 95% confidence intervals. Asterisk (*) denotes $p < 0.05$.

The paradigm yielded a visual bias in the simultaneous flash condition, with a difference of 2.5 degrees between the means of the two conditions. The original study by (Bertelson & Aschersleben, 1998) reported this difference between conditions as differences in location between the left and right staircases of each condition, with the SF condition resulting in a greater mean inter-staircase difference than the NF condition. The SF condition resulted in a mean difference of 4.3 phase difference units, while the NF condition had a mean difference of 1.1 phase difference units. Assuming the distance from the center is half of the difference between the left and right staircases, this results in a mean distance from the center of 2.15 units for the SF condition, and 0.55 units for the NF

condition, which corresponds to a difference of approximately 4 degrees between the two conditions, much larger than found in our experiment.

Experiment 2.1

Because the visual bias in the original study was replicated, the next step was to use this experimental procedure to assess whether the amplitude envelope of the auditory stimuli used would result in differences in this bias. The following experiment tests this for two envelopes, damped and flat, at a duration above the envelope discrimination threshold found in Chapter 2.

1) Methods

a) Subjects

Twelve participants (11 female, 1 male; mean age 19) from the McMaster University Psychology Participant Pool took part in four 1-hour sessions. Participants received $10 per session for a total of $40. All participants signed a consent form and were screened for normal hearing thresholds prior to participating in the experiment. This research was approved by the McMaster University Research Ethics Board.

b) Auditory stimuli

The experiment made use of flat and damped tones with a 2 kHz sinewave and an absolute duration of 83 msec. Flat tones had a 5 msec linear ramp onset and offset, and a sustained amplitude of half the maximum amplitude of the other tones to equate their subjective loudness. Damped tones had a linear ramp onset of 5 msec and an exponential

decay described by the equation $x(t) = e^{(-5t/T)}$ where t is the time and T is the duration (Schlauch, Ries, & DiGiovanni, 2001). The stimuli had A-weighted, fast impulse intensities of 64.0 dB SPL for both envelopes.

c) Conditions

The script manipulated two independent variables: audiovisual condition (synchronous flash, SF versus no flash, NF) and amplitude envelope (flat versus damped). For the synchronous flash condition, a white dot with a 1.5 cm diameter flashed on the center of the screen simultaneously with the presentation of each auditory stimulus. For the no flash condition, a blank screen accompanied the presentation of the stimuli. Each 1-hour session was split into 4 different blocks, each of which had a different combination of the levels of the independent variables: SF-Flat, SF-Damped, NF-Flat, and NF-Damped. Each block was composed of two successive pairs of interleaved staircases. Participants had the possibility to take breaks between the pairs of staircases and blocks and advanced through the different parts of the experiment using the spacebar. Block order was counterbalanced across sessions using a balanced latin square. Order of the days was counterbalanced across participants using a second balanced latin square.

2) Results

Discrimination thresholds, in degrees from the midline, were calculated by obtaining the mean of the last six reversals for each staircase. Table 3 shows descriptive statistics for each envelope and audiovisual condition.

Table 3: Descriptive statistics by envelope and audiovisual condition.

Envelope	**Flat**		**Damped**	
Audiovisual condition	**No Flash**	**Synchronized Flash**	**No Flash**	**Synchronized Flash**
Mean (degrees)	15.4	16.2	15.0	18.2
SD (degrees)	3.35	3.41	3.11	3.67
SEM (degrees)	0.968	0.985	0.898	1.06
Median (degrees)	12.2	13.4	12	14.8
Skewness	1.15	1.08	1.29	0.970

Non-parametric bias-corrected and accelerated bootstrap confidence intervals (95%) were computed using the bcanon function part of the bootstrap package in R (Efron & Tibshirani, 1994). As shown in Figure 4, participants did not demonstrate visual bias, nor were thresholds for the two envelopes different.

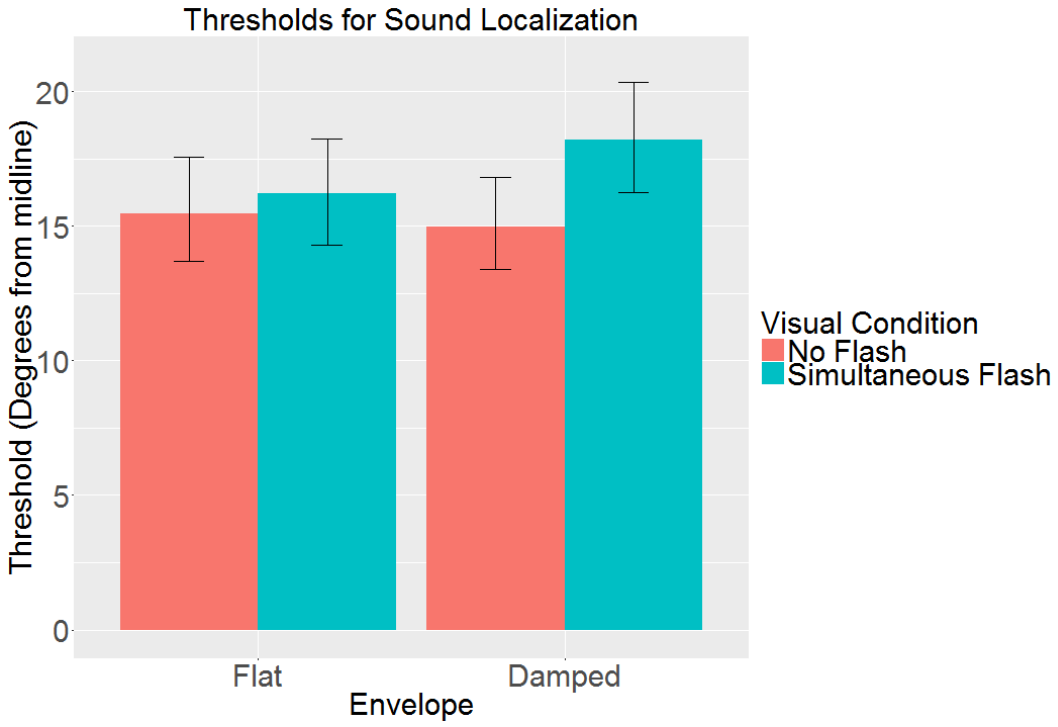

Figure 3: Sound localization thresholds by envelope and audiovisual condition.

Error bars denote bootstrap 95% confidence intervals.

Experiment 2.2

Because there was no visual bias in experiment 2.1, and based on the fact that the auditory stimuli used were much longer than those used in experiment 1, we conducted the experiment again, this time with shorter sounds. The duration used here is sub-threshold for perceptually discriminating envelope differences and serves as a test to determine at which duration the visual bias breaks down using this paradigm.

1) Methods

a) Subjects

Eight participants (4 female, 4 male; mean age 24) from the McMaster University Psychology Participant Pool took part in four 1-hour sessions. Participants received $10 per session and $10 for completing all four sessions for a total of $50. All participants filled a consent form and were screened for normal hearing thresholds prior to participating in the experiment. This research was approved by the McMaster University Research Ethics Board.

b) Auditory stimuli

Stimuli were the same as in experiment 2.1, but had a duration of 32 msec. The A-weighted, fast impulse intensities of the stimuli were of 66.0 dB SPL for the flat tones, and 65.6 dB SPL for the damped tones.

c) Conditions

Conditions and counterbalancing were the same as in experiment 2.1.

2) Results

Discrimination thresholds, in degrees from the midline, were calculated by obtaining the mean of the last six reversals for each staircase. Table 3 shows descriptive statistics for each envelope and audiovisual condition.

Table 4: Summary statistics by amplitude envelope and audiovisual condition.

Envelope	**Flat**		**Damped**	
Audiovisual condition	N**o Flash**	**Synchronized Flash**	**No Flash**	**Synchronized Flash**
Mean (degrees)	13.6	14.6	12.4	12.8
SD (degrees)	2.88	3.11	2.79	2.94
SEM (degrees)	1.02	1.10	0.988	1.04
Median (degrees)	10.6	11.2	6.4	8
Skewness	1.18	1.01	0.944	1.24

Non-parametric bias-corrected and accelerated bootstrap confidence intervals (95%) were computed using the bcanon function part of the bootstrap package in R (Efron & Tibshirani, 1994). As shown in Figure 2, participants did not demonstrate visual bias, nor were thresholds for the two envelopes different.

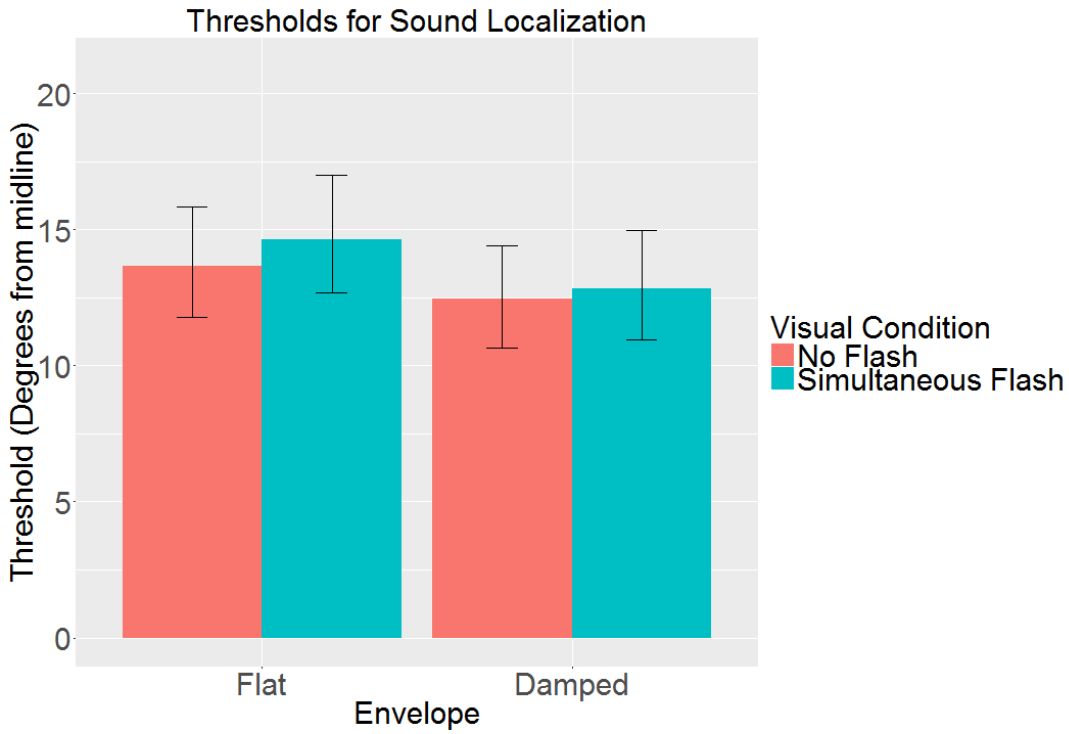

Figure 4: Discrimination thresholds and bootstrap 95% confidence intervals by envelope and audiovisual condition.

3) Discussion

While a visual bias was present in our replication experiment 2.0, none was present in experiments 2.1 and 2.2. These results call into question the possibility of testing the effect of envelope using such a psychophysical method. Despite having very similar intensities, longer stimuli are perceived as louder (Stevens & Hall, 1966), which might have made longer stimuli easier to localize, resulting in a failure to bind them. Because the visual bias was present at 16 msec, but not at 32 or 83 msec, it seems that very short stimulus durations are necessary to provoke ventriloquism using this method, which would explain why stimulus durations of 16 msec and shorter are so common in the ventriloquism literature. Even at 32 msec, the stimuli used fall below the thresholds for envelope recognition found in Chapter 2; as such, even if one could potentially elicit a

visual bias with sounds having different amplitude envelopes and a duration of 16 msec, it is unclear whether the cues provided by envelopes at such a short duration would even be perceptible. Duration constraints using such a framework may limit the magnitude of the effect envelope could have.

Despite replicating the experimental set-up of the original study as closely as possible, the magnitude of the visual bias in the Experiment 2.0 was much smaller than that reported by Bertelson & Aschersleben (1998). The participants in Experiment 2.0 also performed worse than the participants in the original study, with their mean thresholds being 15 degrees away from the center in the NF condition where the participants in the original study were on average about 1 degree from it. Such a large difference may partly be due to the fact that the staircases in the original study were able to cross the midline, whereas those in the present experiments only reached 2.4 degrees from the midline; however, this would not result in the large difference observed here. Even with this difference in the staircases, if the participants in the present experiment were performing as well as the participants in the original study, we would expect their reversals to occur at the location closest and second closest to the midline, resulting in a threshold of 3.6 degrees.

Other methodological differences may also have contributed to a lack of visual bias where others have found one. Rather than present stimuli from an array of possible locations, many studies use three or four possible locations and require participants to indicate from which of these locations they think the sound came (e.g. Slutsky & Recanzone, 2001). Such experimental designs may better capture visual bias, especially

if the magnitude of such a bias is smaller for stimuli devoid of contextual cues. Many studies also present the visual event at different locations throughout the experiment (e.g. Bonath et al., 2014). This strategy may prevent the perceptual system from determining that the auditory and visual events are separate through repeated presentations of the visual stimulus in a static location while the auditory source moves. Another difference which may have influenced the magnitude of the visual bias observed here is the duration of the experiment. To account for the additional conditions introduced by observing the effect of envelope, participants in the current experiments took part in four sessions, each composed of eight pairs of interleaved staircases. Compared to the two sessions composed of six pairs of interleaved staircases used by Bertelson and Aschersleben (1998), this is a marked increase in the time spent doing the task, which may have resulted in participants losing concentration and performing worse at the task.

In terms of determining the effect of amplitude envelope, it may be preferable to take apart complex stimuli in order to determine the relative contribution of each characteristic of stimuli. Chuen & Schutz (2016) did this in a recent study after finding that participants performed worse on an audiovisual temporal order judgement (TOJ) task when the auditory and visual stimuli were matched (sounds and video of the same instrument versus a different one). In order to determine whether the amplitude envelope of the sounds was sufficient to signal unity, they repeated the task using the same envelopes and pure tone carriers rather than the original timbres of the instruments. Using this method, they found that envelope alone was not sufficient to cue the unity of the visual and auditory stimuli.

Most studies looking at spatial ventriloquism using longer stimuli use speech stimuli, as these are more likely to bind (Vatakis & Papadelis, 2014). Future experiments could use speech sounds and visuals to test the effect of envelope, since the amplitude envelope of sounds contributes to our perception of formants in speech. Sounds could be altered using software such as Praat, allowing for a level of control sufficient to ensure internal validity.

Once visual bias is elicited, should an effect of envelope on auditory bias arise, it may be of interest to determine what underlies this effect. Decomposing the sound to isolate binaural and monaural cues could shed more insight into this: is the effect mainly driven by interaural time differences or intensity differences? How do spectral cues from the pinna contribute to the differences in how different envelopes are processed?

Finally, obtaining visual bias would allow for the study of differences in audiovisual integration for atypical populations. As mentioned in Chapter 2, individuals with dyslexia often show deficits in processing of amplitude envelopes. Additionally, individuals with autism demonstrate deficits in audiovisual integration (Brandwein et al., 2012), as do schizophrenic individuals (Williams, Light, Braff, & Ramachandran, 2010). Not only could this provide insight as to the nature of the neurological differences in these populations, but characterizing the differences in performance on audiovisual tasks between neuro-typical and neuro-atypical populations could allow for the creation of unbiased and accessible diagnostic tests.

Chapter 4
Conclusion

As shown in Chapter 2, we can perceive differences in envelopes at very short durations. Because amplitude envelope can change our perception of a wide array of sound properties, this finding has important repercussions for the choice of stimuli in psychophysics experiments. While many studies focus on the effect of amplitude envelope itself, surveys of perceptual research show that studies that do not focus on this characteristic of sound tend not to report the type of envelope of their stimuli. The importance of amplitude envelope should be recognized across the field, such that the envelopes of sounds are reported in the methods of every paper, just like the carrier frequency or duration.

While there is reason to believe that amplitude envelope may affect audiovisual integration, testing this using methods devoid of context may be difficult. Audiovisual illusions such as spatial ventriloquism are difficult to elicit using the pared-down stimuli commonly used in psychophysical paradigms. In Chapter 3, I was able to demonstrate visual bias using sounds with a duration of 16.7 msec, but not for slightly longer durations. This made it impossible to determine the effect of envelope on ventriloquism, as no visual bias was present for the shortest duration above envelope discrimination threshold. It may well be that the most powerful binding occurs in stimuli related by context.

Based on these findings, research bridging psychophysical approaches and auditory scene analysis perspectives is likely to yield interesting findings, highlighting discrepancies between the nature of stimuli frequently employed in studies focused on listening for properties and studies focused on listening for events.

While the inability to elicit visual bias in Chapter 3 precludes the ability to assess my original question, there is much we can gain from looking at these results. The fact that a difference of 16 milliseconds in the duration of the auditory stimuli was enough to break the illusion is a testament to the perceptual system's ability to make accurate sense of multimodal signals. The reason psychophysics studies often need to use such short durations may simply be that it is necessary to deprive the system of information to get a reliable illusory percept. On the other hand, studies using more ecologically valid multimodal stimuli instead rely on these stimuli containing abundant information to signal that the modalities "go together", resulting in strong binding for longer stimuli. The extra information supplied by these stimuli is both low-level (such as cross-modal temporal correlations) and high-level (contextual cues that signal the auditory and visual go together).

Relating this to optimal integration, if the perceptual system is so apt at disentangling multimodal sources, then it must in some way be reducing noise. As mentioned in the previous chapter, the difference in binding between the short pulses and the longer stimuli may be due to a difference in perceived loudness. If that is the case, louder stimuli may provide the system with a less noisy signal, reducing (or in this case, getting rid of) the previously observed visual bias. However, it is impossible to know

from these findings alone whether the perceptual system does this through heuristics or the calculations put forth by the proponents of optimal integration.

Methodologically speaking, researchers interested in listening for properties and listening for events have much to gain from taking the other field into perspective. There is an opportunity to explore the middle ground between the two fields, yielding great insights into perception as a whole, combining the high- and low-level points of view. Deconstructing complex stimuli to determine how stimulus properties lead to unity and binding (e.g. Chuen & Schutz, 2016) is an excellent way to do this.

Some may argue that the goal of psychophysics is to relate the perception of sounds to the structure of the auditory system, for example to develop better signal processing algorithms, and as such, whether or not the sounds used resemble those of the outside world is irrelevant. However, psychophysicists have much to gain besides external validity from keeping the importance of amplitude envelope in mind. Much like our perception of different modalities are integrated in multimodal perception, our perception of the different characteristics of an auditory stimulus cannot be separated. Characterising the contribution of amplitude envelope to our perception of different aspects of sounds, such as duration, timbre, or pitch, and maintaining awareness of this contribution, may help in interpreting the results of experiments, as well as synthesizing the results of many different experiments together. Without the knowledge of what envelopes have been used in different experiments, it is impossible to determine whether discrepancies in envelope can explain discrepancies in the results.

References

Akeroyd, M. A., & Patterson, R. D. (1997). A comparison of detection and discrimination of temporal asymmetry in amplitude modulation. *The Journal of the Acoustical Society of America*, *101*(1), 430–439.

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*(3), 257–262.

Alais, D., Newell, F. N., & Mamassian, P. (2010). Multisensory processing in review: from physiology to behaviour. *Seeing and Perceiving*, *23*(1), 3–38.

Bernstein, L. R. (2001). Auditory processing of interaural timing information: New insights. *Journal of Neuroscience Research*, *66*(6), 1035–1046. http://doi.org/10.1002/jnr.10103

Bertelson, P., & Aschersleben, G. (1998). Automatic visual bias of perceived auditory location. *Psychonomic Bulletin & Review*, *5*(3), 482–489.

Bonath, B., Noesselt, T., Krauel, K., Tyll, S., Tempelmann, C., & Hillyard, S. A. (2014). Audio-visual synchrony modulates the ventriloquist illusion and its neural/spatial representation in the auditory cortex. *Neuroimage*, *98*, 425–434.

Bonath, B., Noesselt, T., Martinez, A., Mishra, J., Schwiecker, K., Heinze, H.-J., & Hillyard, S. A. (2007). Neural basis of the ventriloquist illusion. *Current Biology*, *17*(19), 1697–1703.

Bowers, J. S., & Davis, C. J. (2012). Bayesian just-so stories in psychology and neuroscience. *Psychological Bulletin*, *138*(3), 389–414. http://doi.org/10.1037/a0026450

Brandwein, A. B., Foxe, J. J., Butler, J. S., Russo, N. N., Altschuler, T. S., Gomes, H., & Molholm, S. (2012). The development of multisensory integration in high-functioning autism: high-density electrical mapping and psychophysical measures reveal impairments in the processing of audiovisual inputs. *Cerebral Cortex*, bhs109.

Bregman, A. S. (1994). *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press.

Brown, C. H., Beecher, M. D., Moody, D. B., & Stebbins, W. C. (1980). Localization of noise bands by Old World monkeys. *The Journal of the Acoustical Society of America*, *68*(1), 127–132.

Burr, D., & Alais, D. (2006). Combining visual and auditory information. *Progress in Brain Research*, *155*, 243–258.

Cavonius, C. R., & Robbins, D. O. (1973). Relationships between luminance and visual acuity in the rhesus monkey. *The Journal of Physiology*, *232*(2), 239–246.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America*, *25*(5), 975–979.

Choe, C. S., Welch, R. B., Gilford, R. M., & Juola, J. F. (1975). The "ventriloquist effect": Visual dominance or response bias? *Perception & Psychophysics*, *18*(1), 55–60.

Chuen, L., & Schutz, M. (2016). The unity assumption facilitates cross-modal binding of musical, non-speech stimuli: The role of spectral and amplitude envelope cues. *Attention, Perception, & Psychophysics*, 1–17.

Colavita, F. B. (1974). Human sensory dominance. *Perception & Psychophysics*, *16*(2), 409–412.

Colin, C., Radeau, M., Soquet, A., Dachy, B., & Deltenre, P. (2002). Electrophysiology of spatial scene analysis: the mismatch negativity (MMN) is sensitive to the ventriloquism illusion. *Clinical Neurophysiology*, *113*(4), 507–518.

DiGiovanni, J. J., & Schlauch, R. S. (2007). Mechanisms responsible for differences in perceived duration for rising-intensity and falling-intensity sounds. *Ecological Psychology*, *19*(3), 239–264.

Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC press.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433.

Forrest, T. G., & Green, D. M. (1987). Detection of partially filled gaps in noise and the temporal modulation transfer function. *The Journal of the Acoustical Society of America*, *82*(6), 1933–1943.

Freed, D. J. (1990). Auditory correlates of perceived mallet hardness for a set of recorded percussive sound events. *The Journal of the Acoustical Society of America*, *87*(1), 311–322.

Gaver, W. W. (1993). What in the world do we hear?: An ecological approach to auditory event perception. *Ecological Psychology*, *5*(1), 1–29.

Germann, K. (2016). *The absolute duration threshold of amplitude envelope recognition* (undergraduate thesis). McMaster University, Hamilton, Canada.

Gillard, J., & Schutz, M. (2013). The importance of amplitude envelope: Surveying the temporal structure of sounds in perceptual research. In *Proceedings of the Sound and Music Computing Conference* (pp. 62–68). Citeseer. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.694.9240&rep=rep1&type=pdf

Giordano, B. L., Rocchesso, D., & McAdams, S. (2010). Integration of acoustical information in the perception of impacted sound sources: The role of information accuracy and exploitability. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(2), 462.

Goswami, U., Gerson, D., & Astruc, L. (2010). Amplitude envelope perception, phonology and prosodic sensitivity in children with developmental dyslexia. *Reading and Writing*, *23*(8), 995–1019.

Grassi, M., & Casco, C. (2009). Audiovisual bounce-inducing effect: attention alone does not explain why the discs are bouncing. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(1), 235.

Grassi, M., & Darwin, C. J. (2006). The subjective duration of ramped and damped sounds. *Perception & Psychophysics*, *68*(8), 1382–1392.

Hartmann, W. M. (1978). The effect of amplitude envelope on the pitch of sine wave tones. *The Journal of the Acoustical Society of America*, *63*(4), 1105–1113.

Henning, G. B. (1980). Some observations on the lateralization of complex waveforms. *The Journal of the Acoustical Society of America*, *68*(2), 446–454.

Henning, G. B., & Ashton, J. (1981). The effect of carrier and modulation frequency on lateralization based on interaural phase and interaural group delay. *Hearing Research*, *4*(2), 185–194.

Hudspeth, A. J. (2000). Hearing. In E. R. Kandel, J. H. Schwartz, & T. M. Jessel (Eds.), *Principles of Neural Science* (4th ed.). New York: McGraw-Hill.

Irino, T., & Patterson, R. D. (1996). Temporal asymmetry in the auditory system. *The Journal of the Acoustical Society of America*, *99*(4), 2316–2331.

ISO. (1998). *ISO 389-8. Acoustics: Reference zero for the calibration of audiometric equipment. Part 8: Reference equivalent threshold sound pressure levels for pure tones and circumaural earphones*. International Standards Organization Geneva.

Iverson, P., & Krumhansl, C. L. (1993). Isolating the dynamic attributes of musical timbrea). *The Journal of the Acoustical Society of America*, *94*(5), 2595–2603.

Jackson, C. V. (1953). Visual factors in auditory localization. *Quarterly Journal of Experimental Psychology*, *5*(2), 52–65.

Kietzman, M. L. (1967). Two-pulse measures of temporal resolution as a function of stimulus energy. *JOSA*, *57*(6), 809–813.

King, A. J., & Palmer, A. R. (1985). Integration of visual and auditory information in bimodal neurones in the guinea-pig superior colliculus. *Experimental Brain Research*, *60*(3), 492–500.

Klatzky, R. L., Pai, D. K., & Krotkov, E. P. (2000). Perception of material from contact sounds. *Presence: Teleoperators and Virtual Environments*, *9*(4), 399–410.

Krimphoff, J., McAdams, S., & Winsberg, S. (1994). Caractérisation du timbre des sons complexes. II. Analyses acoustiques et quantification psychophysique. *Le Journal de Physique IV*, *4*(C5), C5–625.

Lee, J. (1994). Amplitude modulation rate discrimination with sinusoidal carriers. *The Journal of the Acoustical Society of America*, *96*(4), 2140–2147.

Lee, J., & Bacon, S. P. (1997). Amplitude modulation depth discrimination of a sinusoidal carrier: Effect of stimulus duration. *The Journal of the Acoustical Society of America*, *101*(6), 3688–3693.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America*, *49*(2B), 467–477.

Martin, F. N., & Champlin, C. A. (2000). Reconsidering the limits of normal hearing. *American Academy of Audiology*, *11*(2), 64–66.

McGrath, M., & Summerfield, Q. (1985). Intermodal timing relations and audio-visual speech recognition by normal-hearing adults. *The Journal of the Acoustical Society of America*, *77*(2), 678–685.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748.

Michalewski, H. J., Starr, A., Nguyen, T. T., Kong, Y.-Y., & Zeng, F.-G. (2005). Auditory temporal processes in normal-hearing individuals and in patients with auditory neuropathy. *Clinical Neurophysiology*, *116*(3), 669–680. http://doi.org/10.1016/j.clinph.2004.09.027

Miller, L. M., & Recanzone, G. H. (2009). Populations of auditory cortical neurons can accurately encode acoustic space across stimulus intensity. *Proceedings of the National Academy of Sciences, 106*(14), 5931–5935.

Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: examining temporal ventriloquism. *Cognitive Brain Research*, *17*(1), 154–163.

Muneaux, M., Ziegler, J. C., Truc, C., Thomson, J., & Goswami, U. (2004). Deficits in beat perception and dyslexia: Evidence from French. *NeuroReport*, *15*(8), 1255–1259.

Narins, P. M., Grabul, D. S., Soma, K. K., Gaucher, P., & Hödl, W. (2005). Cross-modal integration in a dart-poison frog. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(7), 2425–2429.

Neuhoff, J. G. (1998). Perceptual bias for rising tones. *Nature*, *395*(6698), 123–124.

Neuhoff, J. G. (2001). An adaptive bias in the perception of looming auditory motion. *Ecological Psychology*, *13*(2), 87–110.

Nuetzel, J. M., & Hafter, E. R. (1976). Lateralization of complex waveforms: Effects of fine structure, amplitude, and duration. *The Journal of the Acoustical Society of America*, *60*(6), 1339–1346.

Pasquini, E. S., Corriveau, K. H., & Goswami, U. (2007). Auditory processing of amplitude envelope rise time in adults diagnosed with developmental dyslexia. *Scientific Studies of Reading*, *11*(3), 259–286.

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1), 8–13.

Phan, M. L., Schendel, K. L., Recanzone, G. H., & Robertson, L. C. (2000). Auditory and visual spatial localization deficits following bilateral parietal lobe lesions in a patient with Balint's syndrome. *Journal of Cognitive Neuroscience*, *12*(4), 583–600.

Rayleigh, Lord. (1907). XII. On our perception of sound direction. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *13*(74), 214–232.

Recanzone, G. H. (2003). Auditory influences on visual temporal rate perception. *Journal of Neurophysiology*, *89*(2), 1078–1093.

Recanzone, G. H., Makhamra, S. D., & Guard, D. C. (1998). Comparison of relative and absolute sound localization ability in humans. *The Journal of the Acoustical Society of America*, *103*(2), 1085–1097.

Richardson, U., Thomson, J. M., Scott, S. K., & Goswami, U. (2004). Auditory processing skills and phonological representation in dyslexic children. *Dyslexia*, *10*(3), 215–233.

Ries, D. T., Schlauch, R. S., & DiGiovanni, J. J. (2008). The role of temporal-masking patterns in the determination of subjective duration and loudness for ramped and damped sounds. *The Journal of the Acoustical Society of America*, *124*(6), 3772–3783.

Sabin, A. T., Macpherson, E. A., & Middlebrooks, J. C. (2005). Human sound localization at near-threshold levels. *Hearing Research*, *199*(1), 124–134.

Sanabria, D., Soto-Faraco, S., Chan, J. S., & Spence, C. (2004). When does visual perceptual grouping affect multisensory integration? *Cognitive, Affective, & Behavioral Neuroscience*, *4*(2), 218–229.

Sato, Y., Toyoizumi, T., & Aihara, K. (2007). Bayesian inference explains perception of unity and ventriloquism aftereffect: identification of common sources of audiovisual stimuli. *Neural Computation*, *19*(12), 3335–3355.

Schlauch, R. S., Ries, D. T., & DiGiovanni, J. J. (2001). Duration discrimination and subjective duration for ramped and damped sounds. *The Journal of the Acoustical Society of America*, *109*(6), 2880–2887.

Schutz, M., & Kubovy, M. (2009). Causality and cross-modal integration. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(6), 1791.

Schutz, M., & Lipscomb, S. (2007). Hearing gestures, seeing music: Vision influences perceived tone duration. *Perception*, *36*(6), 888.

Schutz, M., & Vaisberg, J. M. (2014). Surveying the temporal structure of sounds used in Music Perception. *Music Perception: An Interdisciplinary Journal*, *31*(3), 288–296.

Shams, L., Ma, W. J., & Beierholm, U. (2005). Sound-induced flash illusion as an optimal percept. *Neuroreport*, *16*(17), 1923–1927.

Sheft, S., & Yost, W. A. (1990). Temporal integration in amplitude modulation detection. *The Journal of the Acoustical Society of America*, *88*(2), 796–805.

Slutsky, D. A., & Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect. *Neuroreport*, *12*(1), 7–10.

Spence, C. (2007). Audiovisual multisensory integration. *Acoustical Science and Technology*, *28*(2), 61–70. http://doi.org/10.1250/ast.28.61

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, *73*(4), 971–995. http://doi.org/10.3758/s13414-010-0073-7

Stecker, G. C., & Hafter, E. R. (2000). An effect of temporal asymmetry on loudness. *The Journal of the Acoustical Society of America*, *107*(6), 3358–3368.

Stekelenburg, J. J., Vroomen, J., & de Gelder, B. (2004). Illusory sound shifts induced by the ventriloquist illusion evoke the mismatch negativity. *Neuroscience Letters*, *357*(3), 163–166.

Stevens, J. C., & Hall, J. W. (1966). Brightness and loudness as functions of stimulus duration. *Perception & Psychophysics*, *1*(9), 319–327.

Sugita, Y., & Suzuki, Y. (2003). Audiovisual perception: Implicit estimation of sound-arrival time. *Nature*, *421*(6926), 911–911. http://doi.org/10.1038/421911a

Su, T.-I. K., & Recanzone, G. H. (2001). Differential effect of near-threshold stimulus intensities on sound localization performance in azimuth and elevation in normal human subjects. *JARO-Journal of the Association for Research in Otolaryngology*, *2*(3), 246–256.

Vallet, G. T., Shore, D. I., & Schutz, M. (2014). Exploring the role of the amplitude envelope in duration estimation. *Perception*, *43*(7), 616–630.

Van Heuven, V., & Van Den Broecke, M. P. R. (1979). Auditory discrimination of rise and decay times in tone and noise bursts. *The Journal of the Acoustical Society of America*, *66*(5), 1308–1315.

Vatakis, A., & Papadelis, G. (2014). 20 The Research on Audiovisual Perception of Temporal Order and the Processing of Musical Temporal Patterns: Associations, Pitfalls, and Future Directions. *Subjective Time: The Philosophy, Psychology, and Neuroscience of Temporality*, 409.

Vatakis, A., & Spence, C. (2007). Crossmodal binding: Evaluating the "unity assumption" using audiovisual speech stimuli. *Perception & Psychophysics*, *69*(5), 744–756.

Viemeister, N. F. (1970). Intensity discrilnination: Performance in three paradigms. *Perception & Psychophysics*, *8*(6), 417–419.

Viemeister, N. F. (1979). Temporal modulation transfer functions based upon modulation thresholds. *The Journal of the Acoustical Society of America*, *66*(5), 1364–1380.

Viemeister, N. F., & Wakefield, G. H. (1991). Temporal integration and multiple looks. *The Journal of the Acoustical Society of America*, *90*(2), 858–865.

Warren, W. H., & Verbrugge, R. R. (1984). Auditory perception of breaking and bouncing events: a case study in ecological acoustics. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(5), 704.

Welch, R. B. (1999). Meaning, attention, and the "unity assumption" in the intersensory bias of spatial and temporal perceptions. *Advances in Psychology*, *129*, 371–387.

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, *88*(3), 638.

Williams, L. E., Light, G. A., Braff, D. L., & Ramachandran, V. S. (2010). Reduced multisensory integration in patients with schizophrenia on a target detection task. *Neuropsychologia*, *48*(10), 3128–3136.

Zampini, M., Shore, D. I., & Spence, C. (2003). Audiovisual temporal order judgments. *Experimental Brain Research*, *152*(2), 198–210.

Appendix A

Informal survey of auditory stimulus durations in spatial ventriloquism studies

Articles using speech stimuli were not included.

Authors	Journal	Year	Duration(s) (msec)
Morein-Zamir et al.	Cognitive Brain Research	2003	5
Slutsky & Recanzone	Neuroreport	2000	200
Vroomen & De Gelder	JEP:HPP	2004	16.7
Bertelson & Aschersleben	Intl. Journal of Philosophy	2003	15
Spence & Driver	Neuroreport	2000	100
Radeau & Bertelson	Perception & Psychophysics	1976	50
Vroomen & Keetels	JEP:HPP	2006	5
Parise & Spence	Neuroscience Letters	2008	5
Keetels et al.	Experimental Brain Research	2007	3
Bertelson et al.	Neuropsychologia	2000	120
Bischoff et al.	Neuropsychologia	2007	120
Stekelenburg & Vroomen	Neuroreport	2005	16.7
Bertini et al.	European Journal of Neuroscience	2010	100
Hartcher-O'Brien & Alais	JEP:HPP	2011	10
Cook & Van Valkenburg	Perception	2009	1250