A theoretical framework for modeling pedestrians crossing behavior along a trip

Eleonora Papadimitriou[1], George Yannis[2], John Golias[3]

Abstract

Explaining pedestrians crossing behavior along entire trips may contribute towards more efficient and targeted planning of pedestrian facilities in urban areas and more accurate consideration of pedestrian safety. Although existing research on pedestrians crossing behavior is extensive, most related studies examine pedestrians crossing decisions at local level and focus on a particular set of determinants (roadway environment, traffic conditions, or human factors). On the contrary, crossing behavior along entire trips is seldom explored. This paper presents a theoretical framework for modeling pedestrians crossing behavior along a trip, addressing a large part of the difficulties involved in collecting the necessary data and setting up a modeling

---

[1] Research Associate, Department of Transportation Planning and Engineering, National Technical University of Athens
Address: 5, Iroon Polytechniou str., GR-15773 Athens, Greece
tel.: +30 210 7721380, fax: +30 210 7721454
e-mail: nopapadi@central.ntua.gr

[2] Assistant Professor, Department of Transportation Planning and Engineering, National Technical University of Athens
Address: 5, Iroon Polytechniou str., GR-15773 Athens, Greece
e-mail: geyannis@central.ntua.gr

[3] Professor, Department of Transportation Planning and Engineering, National Technical University of Athens
Address: 5, Iroon Polytechniou str., GR-15773 Athens, Greece
e-mail: igolias@central.ntua.gr

framework. First, a topological approach of pedestrian trip characteristics and crossing decisions is proposed, allowing to consider distinct patterns of crossing behavior along a trip. Moreover, specific techniques from the family of discrete choice models are proposed for determining the number and location of pedestrians' crossings, accounting for the hierarchical and dynamic nature of pedestrians decisions along a trip. Finally, a field survey method is presented, allowing to collect detailed information about pedestrians crossing decisions along urban trips, including data on roadway, traffic and individual characteristics, as well as the interactions between pedestrians and motorists. Preliminary results from a pilot implementation of the proposed framework are promising and a full scale application for testing and validation is in progress.

Key-words: pedestrians; behavior; models.

Introduction

Explaining and forecasting pedestrians crossing behavior along a trip is an important yet complex topic, for which limited contribution is available in the literature, mainly due to obvious difficulties involved in defining an appropriate framework for analysis, collecting a considerable amount of data and selecting efficient modeling techniques. However, the knowledge on pedestrians crossing decisions during urban trips may assist towards more accurate estimation of pedestrian's road accident risk and more targeted planning of pedestrian facilities in urban areas.

Existing research results on pedestrian crossing behavior in urban areas range from analyses of pedestrians decisions on road crossing location (Chu et al. 2004; Baltes & Chu 2002), analyses of pedestrians traffic gap acceptance (Hamed 2001; Das et al. 2005), and pedestrians compliance to traffic rules (Sisiopiku & Akin 2003; Yang et al. 2006), to analyses of the effect of traffic engineering measures (Van Houten & Malenfant 1992; Nee & Hallenberg 2003; Keegan & O'Mahony 2003; Hakkert et al. 2002) or traffic control measures (Carsten et al. 1998; Hubbard et al. 2009) on pedestrians behavior.

Most of all these analyses examine crossing behavior at local level, whereas crossing patterns along entire trips are seldom explored. Moreover, they focus on a particular set of determinants of pedestrian behavior. A thorough review of (mostly earlier) studies dealing with pedestrians crossing behavior can be found in Ishaque & Noland (2007). For an exhaustive review and assessment of existing pedestrian crossing behavior models the reader is referred to Papadimitriou et al. (2009), where four main issues are outlined as regards the needs for further research on the topic: the need for analyses at trip level, the need for more explanatory approaches, the need for flexible disaggregate modeling techniques and the need for more extensive data collection schemes.

In particular, the analysis of pedestrian behavior along entire trips requires different and more extensive data collection processes, for obtaining detailed and reliable data. So far, most researchers use data from local level video recordings, that can not be used for the purposes of analyses at trip level (Bierlaire et al. 2003). Other researchers use interview results (Hine 1996; Fitzpatrick et al. 2004), which often suffer from the limitations of self-reporting.

However, the main difficulties that need to be addressed for modeling pedestrian crossing behavior along a trip are conceptual and methodological ones. Pedestrian trips do not take place along strictly fixed corridors of the road environment, and pedestrians themselves are subject to looser traffic rules, in relation to motorists, allowing them more opportunity for individual decision-making. Moreover, pedestrians may interact not only with the roadway and traffic environment, but also with other pedestrians. Finally, pedestrians are adjustable in terms of travel and activity strategy. Consequently, the analysis of pedestrian trips in urban areas and the related crossing decisions can not be systematically tackled in a straightforward way.

Consequently, appropriate modeling techniques need to be identified, so that the complex and dynamic nature of pedestrian crossing decisions along trips may be captured. In this framework, stochastic models of pedestrian movement are not considered to be very promising, as they do not allow for identification of explanatory effects (Papadimitriou et al. 2009). On the other hand, discrete choice models are proposed by several authors as a most appropriate technique for describing pedestrians behavior (Chu et al. 2004; Antonini et al. 2006; Yannis et al. 2007), however an appropriate framework for road crossings along a trip still needs to be defined.

The objective of this paper is to present a theoretical framework for modeling pedestrians crossing behavior along a trip, addressing a large part of the issues and difficulties mentioned above. The crossing behaviour examined concerns the choice of a crossing location among the available alternatives along a trip. It is noted that the need for considering the complex interaction between route choice and crossing behaviour, underlined in recent research (Papadimitriou et al. 2009) is beyond the scope of this paper.

The paper starts with a topological approach of pedestrian crossings, allowing to consider distinct patterns of crossing behavior along a trip. Moreover, techniques from the family of discrete choice models are proposed for determining the location of pedestrians' crossings. Finally, a field survey method is presented, allowing to collect detailed data on pedestrians crossing decisions along urban trips. Preliminary results from a pilot implementation of the proposed framework are presented and some engineering implications of such modeling results are discussed in terms of pedestrian infrastructure design and traffic control.

Conceptual framework

Topology of pedestrian trips and crossings

A first problem that needs to be addressed for modeling pedestrians crossing behavior along a trip is to elaborate a conceptual framework for considering pedestrians crossing decisions in a systematic way. Once this is achieved, it can be attempted to localize, count and discriminate the different types of crossings (crossing a road at a junction or at mid-block, on or out of a pedestrian crosswalk etc.). On that purpose, a topological analysis framework is opted for.

In particular, the road network of a pedestrian trip can be typically represented by a graph, where links (edges) correspond to roads traveled and nodes (vertices) correspond to junctions encountered (Figure 1a). The graph links can be further distinguished into those corresponding to the road links along which the pedestrian walks (these are shown with a continuous line in Figure 1a and will be from now on referred to as primary links) and those corresponding to other arms of the junctions through which the pedestrian walks (these are shown with a dotted line in Figure 1a and will be from now on referred to as secondary links). It is therefore noted that the classification of road links into primary and secondary depends on the form of the trip (origin, destination, path). This analysis focuses on the topological properties of the graph, i.e. those properties that are maintained in continuous (homeomorphic) transformations of the graph (e.g. stretching, bending etc.) and therefore finer properties (e.g. link length, grid degree etc.) are not equally important.

Most of the components of a pedestrian path (e.g. origin, destination, walking movements) are not situated on the graph itself, as these do not take place on the actual road network, but on the adjacent space (e.g. sidewalks). On the other hand, road crossing movements intersect the graph, either on primary or on secondary links (see Figure 1b). Obviously, if both the origin and destination of the pedestrians lie on the same 'side' of the graph, the pedestrian's path will not necessarily intersect a primary link of the graph, whereas if the origin and destination are on different 'sides' of the graph, the pedestrian path will definitely intersect a primary link of the graph. It is noted that in either case, the pedestrian path may intersect with secondary links of the graph.

These intuitive ideas have a well established topological basis, mainly with respect to the Jordan curve theorem for closed curves and its extensions for other topological objects (Guggenheimer 1977; Moore 2008), according to which, a curve divides the plane space into two distinct sets, an 'interior' and an 'exterior' one and any path from one set to the other intersects the curve. Furthermore, a path connecting two points from the same set intersects the curve an even number of times, whereas a path connecting two points from different sets intersects the curve an odd number of times (see Figure 2). The consideration of pedestrian networks as topological graphs allows for analyzing pedestrian paths and the related crossings in such a topological context.

Figure 2 to be inserted here

In particular, for any pedestrian network as a topological graph, an 'interior' and an 'exterior' set can be defined (e.g. in relation to the convexity of the graph), so that the origin and destination of the pedestrian path can be localized accordingly. Consequently, an odd or even number of pedestrian crossings of primary links of the graph will be expected. As mentioned above, crossings of the secondary links of the graph may also be observed. The distinction of 'primary' and 'secondary' crossings of the trip graph has another useful implication. It can be shown that primary crossings are probabilistic, whereas secondary crossings are deterministic (Lassarre et al. 2007).

More specifically, a pedestrian path may differ in terms of the point of intersection with the primary graph, given that the pedestrian may choose between several options (junctions, crosswalks etc) along the road links of the trip. On the other hand, each secondary crossing will take place at a given location and this depends on the occurrence of primary crossings. For example, in Figure 1b, for a particular path (b), a primary crossing takes place across the first link of the graph; therefore a secondary crossing takes place around the second node of the graph (and within the exterior set). However, in an alternative path (a) for the same origin-destination configuration, the primary crossing takes place across the second link of the graph, and the secondary crossing around the second node necessarily takes place within the interior set of the graph. This simple observation allows to limit the analysis of pedestrians crossing behavior to primary crossings only, since the location of secondary crossings can be fully determined if the locations of primary crossings are known. It is noted that this classification aims to serve the analysis purposes and not imply an increased importance of a certain type of crossings.

Figure 1 to be inserted here

Finally, the exploitation of the graph topology (i.e. interior and exterior set) may assist not only in determining the number of primary crossings (odd or even) but also in localizing those crossings on specific links of the graph. For example, in Figure 1b, where the trip origin lies on the exterior set of the graph and the destination lies on the exterior set of the graph, a primary crossing may take place either on the first (path (b)) or the second (path (a)) primary link of the graph, together with a secondary crossing at the second node; however, it is unlikely that a primary crossing would take place on the third primary link of the graph, as this would not only imply a detour, but also add two 'unnecessary' secondary crossings at the third node. Accordingly, one may identify specific crossing scenarios for each case, in which primary crossings may take place only on particular links of the graph. This may eventually allow the definition of pedestrians' choice set as regards road crossing location, as will be shown in the next section.

Pedestrian crossing scenarios

The above topological approach of pedestrian trips allows the definition of specific road crossing scenarios for each trip. In particular, one may determine an expected total number of primary crossings and localize each one of them on specific parts of the graph. This translates into the definition of a crossing choice set of alternative locations (road links) for each one of the primary crossings expected. As mentioned earlier, each road link traveled by pedestrians may include a number of alternative crossing locations to be considered within the pedestrian's decision making process. Therefore, determining the road links (graph links) to be included or not in this decision making process is a critical first step in building a choice model.

In order to address this process in a systematic way, four basic families of graphs are initially considered, as shown in Figure 3, ranging from simple linear shape, to "gamma" shape of two links with one bend, and then to "pi" shape of three links with two bends of the same direction, and to "sigma" shape of three links with two bends of different directions. It is expected that most pedestrian trips may be topologically represented by one of these graphs, whereas more complex trips may be decomposed into combinations of these basic graphs. Moreover, it is also possible to extend these graphs for more than one link per direction. In each case, an interior and an exterior set of the space around the graph may be defined, as in Figure 3.

Figure 3 to be inserted here

In order to come up with a finite number of crossing scenarios, it is necessary to exclude those scenarios that result in unrealistically large detours and increased or unnecessary crossings. A practical case-specific example was provided in the previous section. The general principle for excluding scenarios can be outlined as follows: primary crossings resulting in a detour so that an opposite secondary crossing will be required to reach the destination are not considered. In other words, scenarios including pairwise opposite crossings are not considered. Such a not applicable scenario is demonstrated in Figure 4, for a gamma-shaped trip graph with origin and destination lying on the interior set of the graph. On the other hand, Figure 5 shows the applicable scenarios and related crossing choice sets for different origins and destinations around gamma-shaped graphs.

Figure 4 to be inserted here

Figure 5 to be inserted here

On the basis of these principles, nineteen scenarios of pedestrians crossing behavior along a trip can be defined. These are summarized in Table 1. Moreover, as can be seen in Table 1, in some scenarios, the links of the choice set may or may not include a single roadway, as is the case e.g. in 'Pi'-shaped trips.

Through these scenarios, pedestrians primary crossing choices along a trip are identified. Consequently, a choice model can be defined for each primary crossing of each scenario, in which a pedestrian chooses a crossing location among a finite set of alternatives. In the following sections, specific modeling structures and techniques are proposed.

Table 1 to be inserted here

Modeling framework

In the present research, discrete choice models (Ben-Akiva & Lerman 1985) are selected as a most promising technique for explaining pedestrians crossing behavior along a trip. First of all, they allow for disaggregate (microscopic) analysis having pedestrians as units. Moreover, they allow for the identification of both systematic and random effects. Moreover, a broad range of techniques is available, including ordered, nested or crossed models. Finally, the dynamics of decision making processes can also be examined.

A finite number of primary crossing options are considered for each trip, according to the proposed crossing scenarios. A probabilistic choice is then involved in determining the location of each primary crossing from the alternatives of the choice set. In particular, each choice set is a sequence of road links, lying on the same or on different directions, each road link including two basic crossing options i.e. junction area and mid-block area. Therefore, the number of actual primary crossing alternatives in each scenario is equal to twice the number of road links in the choice set, i.e. (n) junction areas and (n) mid-block areas for (n) road links in the choice set.

A utility function is associated with each crossing alternative (i), as follows:

$$U_{in} = V_{in} + \varepsilon_{in} \qquad (1)$$

Where $V_{in} = \beta' X_{in}$ is the systematic (deterministic) part of the utility, and εi_n ~ extreme value (0, μ), is the random part of the utility for each alternative, so that for the entire choice set ε_n ~ logistic (0, μ).

However, in order to model the decision making process, it is necessary to make certain assumptions about the structure and nature of this process. In this research, two types of decision making process are proposed, a sequential process and a hierarchical process. It is noted that, it is likely that neither of the proposed processes fully reflects the intrinsic (often subconscious) real-time decision making process of pedestrians. However they are considered to adequately represent the general observed decision making process for analysis purposes.

Hierarchical modeling

The hierarchical process of modeling the location of each primary crossing along pedestrian trips is based on the assumption that the pedestrian considers the entire set of alternatives at the same time. However, these alternatives are not independent. For example, for each primary crossing to be made, pedestrians may first select a road link and then a specific location, either at junction or at mid-block within that link. Consequently, a choice set of road links C_m = {link 1, link 2, .., link m} is initially considered (marginal choice), and then a choice set of crossing options C_i = {Junction, Mid-block} for each road link (conditional choice).

This decision making process may be represented by a Nested Logit Model (NL) according to which, each pedestrian faces a choice within a series of road links, and a subsequent choice of "junction or mid-block" within each road link (Figure 6). In this case, the choice probability would be estimated as:

$$P(i\,|m)\,P(m\mid C) = \frac{e^{\mu_m V_i}}{\sum\limits_{j\in C} e^{\mu_m V_j}} \frac{e^{\mu V_m}}{\sum\limits_{k} e^{\mu V_k}} \qquad (2)$$

Figure 6 to be inserted here

However, in the NL model, it is assumed that pedestrians assess the utilities of the junction and mid-block locations of each link, without considering those of other links i.e. it is assumed that the choice of a road link precedes the choice of a particular crossing location within the link. In practice, the choice of road link and the choice of particular crossing location are most likely interrelated.

More specifically, it can be assumed that each pedestrian faces two parallel choices, one of choosing an appropriate road link C_m = {link 1, link 2, .., link m} and one of choosing between the available junctions or mid-block locations C_i = {Junction 1, Mid-block 1, …, Junction m, Mid-block m}. These two choice sets are combined and may be represented by a cross-nested logit (CNL) model structure (Daly & Bierlaire 2004; Bierlaire 2006), as shown in Figure 7. In this case, the choice probability is:

$$P(i|C) = \sum_{m} P(i|m)P(m|C) = \sum_{m=1}^{m} \frac{\left[\sum_{j\in C} \alpha_{jm}^{\mu_m/\mu} e^{\mu_m V_j}\right]^{\frac{\mu}{\mu_m}}}{\sum_{n=1}^{M}\left(\left[\sum_{j\in C} \alpha_{jn}^{\mu_n/\mu} e^{\mu_n V_j}\right]\right)^{\frac{\mu}{\mu_n}}} \frac{\alpha_{im}^{\mu_m/\mu} e^{\mu_m V_i}}{\sum_{j\in C} \alpha_{jm}^{\mu_m/\mu} e^{\mu V_j}} \qquad (3)$$

Where α_{im} is the weight / membership factor, indicating that each alternative belongs to more than one "nests" with $\sum_{m} \alpha_i = 1$.

Figure 7 to be inserted here

<u>Sequential modeling</u>

The above hierarchical choice assumption is rather restrictive and a more relaxed approach would be worth examining as well. In this case, an alternative assumption would be that a pedestrian examines the available choice set on a sequential, stepwise basis. Therefore, a separate crossing decision may be associated to each link of the choice set.

More specifically, in a sequential modeling approach, a choice set of three alternatives C={Junction, Mid-block, No crossing} is considered for each road link, which corresponds to a multinomial choice model (MNL). The decision making process is therefore described by a sequence of MNL models, as shown in Figure 8. In such a context, the choice of the "no crossing" option in the first MNL implies the passage to the next one, and so on, until one of the "junction or mid-block" options is chosen. Consequently, the remaining choice set is never considered by the pedestrian.

Figure 8 to be inserted here

On each road link, the choice probability is given by:

$$P(i\,|C) = \frac{e^{\mu V_i}}{\sum_{j \in C} e^{\mu V_j}} \qquad (4)$$

Certainly, these sequential choices of the same individuals can not be considered as independent. In fact, two types of dependence may be involved:

- Individual-specific heterogeneity (Wooldridge 2004), i.e. random variation resulting from the fact that these choices are repeated observations by the different individuals (panel effects).
- State dependence (Honore & Kyriazidou 2000)*)*, due to the fact that each choice is made on different states of the same process, and thus the choice in the previous state may affect the choice of the current state, as in a typical 1st order Markov process.

In this case, two types of extensions may need to be incorporated in the utility function of the multinomial model, which is now considered to apply to each state T=t (a separate state corresponding to each link) and takes the following form:

$$U_{int} = \beta X_{nt} + \gamma y_{n,t-1} + \alpha_n + \varepsilon_{int} \qquad (5)$$

Where $y_{n,t-1}$ is the choice made in the previous state T= t-1 (state dependence),
α_n the unobserved heterogeneity, which may be fixed or random $\sim N(0, \sigma^2)$,
and ε_{nt} the random utility.

Data collection framework

In this research, a field survey design is proposed, allowing for the collection of detailed data on pedestrian crossings along a trip and on the road and traffic environment, as well as on the individual characteristics of the pedestrian, for the development of the proposed models.

The survey is based on following pedestrians without their knowledge at different times of the day (morning, evening, off peak hours) in real traffic conditions. The pedestrians are randomly selected from a fixed origin (e.g. metro or bus station) to their first destination (possibly excluding stops at shop windows, kiosks, news stands etc.), and their behavior and the road / traffic environment is video-recorded for the entire trip. Certainly, in order to preserve the pedestrian's privacy as much as possible, the video recording is anonymous, the specific address of the destination is not noted and the individual's face is not captured.

The proposed method allows for an important number of variables to be collected and coded, despite the lack of detailed information on pedestrians' perception, attitudes and preferences. Moreover, important detailed behavioral information, such as the exact number and location of crossings, their characteristics and the characteristics of the alternative crossing options, the traffic gaps accepted at each crossing, the interactions with motorists or with other pedestrians, and the pedestrians compliance to the traffic rules, can be made available for each pedestrian / trip through this method.

More specifically, the variables to be collected can be classified in four categories, namely pedestrians variables (I variables), trip variables (R variables), road link variables (L variables) and primary crossings variables

(Cp variables). These are summarized in Table 2. It can be seen that most of the variables are directly observable while following a pedestrian, except from a limited number of variables, which require a degree of personal judgment (e.g. whether the pedestrian appears to be in a hurry, whether the destination was fixed or occasional etc.).

Table 2 to be inserted here

For each trip, detailed information concerning each road link is recorded; then, a choice indicator for primary crossings is assigned to each link, indicating on which link a primary crossing took place. For each primary crossing, several variables are also recorded. It is also noted that the related secondary crossings are also noted, and can be associated to the end of each primary road link.

Results

Pilot models development

A pilot implementation of the proposed survey design was already carried out, and a total of 200 observations (pedestrian trips) were collected and coded. The survey was carried out around two metro stations of the Athens city centre, the Evangelismos station area (mostly commercial and recreational activities) and the Panormou station area (mostly residential area with few commercial activities). It is noted that the choice of metro stations as pedestrian trip origins may have implications as per the type of trips recorded, and this should be taken into account in the interpretation of the modeling results.

The main difficulties encountered during the pilot field survey concerned the achievement of a successful video recording in motion, without disturbing the pedestrian followed or violating his or her privacy. The pilot survey was proved to be promising as regards the feasibility and effectiveness of the approach. Moreover, the data coding process did not present considerable difficulties.

These 200 trips include more than 800 road links, where 226 primary crossings took place in total, out of which only 74 were at junction. Initially, 56 linear trips were obtained, 75 Gamma trips, 10 pi trips and 40 sigma trips. Moreover, 19 trips presented more complex shape than the four basic shapes considered in this research and were decomposed into combinations of those basic shapes. Finally, only 2 out of the 220 trips eventually obtained could not be classified in one of the proposed scenarios, mainly because these included pairwise opposite crossings. It is also noted that 63 trips belonged to scenarios of zero primary crossings and were not included in the modeling process. Moreover, 58 trips included more than 1 primary crossings (34 trips with 2, 21 trips with 3 and 3 trips with 5 primary crossings); in each case, the proposed scenarios of Table 1 - or combinations of them - were applied in order to determine the choice set of each crossing, defined as a subset of road links of the pedestrian trip

On the basis of this data, pilot model fitting was carried out, in order to assess the feasibility and appropriateness of the proposed model structures. Therefore, the models presented below only include a limited number of key variables. These include the pedestrian's gender (1: male, 0: female), the presence of an attractor, the presence of a marked crosswalk, the road side parking (yes / no), the total trip length and the position of each road link within the trip (lsn1: first link, lsn23: second or third link, lsn45: fourth or fifth link). All models were fitted by using the Biogeme v1.7 dedicated software package for discrete choice models (Bierlaire 2003; Bierlaire 2008).

It is noted that 142 of the choice sets resulting from the application of the scenarios included only one road link. These were handled separately as binary choice models for the "1: junction, 0: mid-block" decision. Indicative results are presented in Table 3. It can be seen that men are more likely to cross at junction than women. Moreover, the higher the serial number of each road link within the trip, the higher the probability of crossing at junction. On the contrary, roadside parking appears to favor crossings at mid-block, confirming previous studies, according to which pedestrians may exploit parked vehicles by waiting on the pavement, reducing thus crossing distance. Moreover, the higher the total trip length, the higher the probability of crossing at mid-block. Finally, the presence of an attractor across the street increases the probability of mid-block crossing. It is noted that, given that all variables are binary ones, their parameter estimates reflect the relative importance of their effect on the crossing probability.

Table 3 to be inserted here

Concerning the choice sets including multiple links, the sequential model was tested first. Given the relatively small dataset, the use of two control variables instead of a typical state dependence term was opted for, these two variables indicating whether the pedestrian has already skipped one or more crossing opportunities on one or more previous links (variables "not previous" and "not previous2"). As shown in Table 4, only the first one is significant, suggesting that if no crossing took place on the previous link, there is increased probability of crossing on the current link. Random heterogeneity was tested as an additional, normally distributed error term, but was found to be non significant, suggesting that differences among individuals are not important in this dataset. As regards the other explanatory variables tested, their effects are in accordance to the ones mentioned above. It is interesting to note that total trip length (fitted as alternative-specific) appears to increase the probability of not crossing for all links, which may be considered as an additional control variable for sequential nature of the decision making process (i.e. pedestrians tend to postpone crossing in longer trips).

Table 4 to be inserted here

Finally, some results of the implementation of the hierarchical approach are presented in Table 5. The related 75 choice sets included up to 10 road links, however all crossing choices were found within the 5 first links. In this case, it is important to keep in mind that the results are conditional on the availability of alternatives, i.e. the utility of each alternative is calculated only for the choice sets that include this alternative. Given the small dataset and the limited number of choice sets including more than 3 road links, a reduced hierarchical structure was tested. In this structure, two nests are considered, one referring to the first link of the choice set, and one referring to all (i.e. any of) the following links of the choice set. Each nest then includes one junction and one mid-block alternative. A cross-nested logit model with fixed membership factors was tested. The membership factors were fixed so that the options of the second nest belong by 50% to the first nest as well (i.e. assuming that the decision of crossing at junction or at mid-block later on during the trip is considered by 50% while being on the first link of the trip).

The pilot results reveal that the parameter of nest 2 is not significantly different from one, suggesting that a cross-nested structure may not be necessary after all. Moreover, the longer the total trip length, the higher the probability of postponing the primary crossing towards the end of the trip. Men appear to be more likely to cross at the first junction available, rather than anywhere else. Finally, the presence of marked crosswalk at junction on the first road link increases the probability of that option being chosen.

Table 5 to be inserted here

The above results can be considered to confirm the feasibility and usefulness of both proposed approaches. However, the effects identified, although statistically significant, can not be further validated at this stage. Extensive testing will be required and is in progress for the identification of all explanatory effects. Furthermore, a lot of additional effort is required for determining the optimal utility functions. Nevertheless, by the different model structures, one may obtain some initial insight of some aspects of the pedestrian decision making process (e.g. the tendency to cross earlier or later in the trip, the effect of rejecting previous crossing opportunities etc.). In the next section, the engineering implications of such modeling results are discussed.

Engineering implications

In general, the objective of the analysis of pedestrians crossing behaviour along urban trips is the understanding of pedestrian decision making and the identification of related determinants, in terms of the role of the infrastructure, the traffic conditions and the characteristics of the pedestrians. The development of models aims to capture the way pedestrians interact with the road and traffic environment, as well as the way pedestrians respond to changes in that environment. Eventually, it should enable testing of alternative engineering or traffic control strategies as per the behaviour of pedestrians, and estimating the related level of service and road safety of pedestrians.

The management of pedestrian flows along entire trips is a most challenging question within this context, for ensuring an uninterrupted and self-explaining walking environment, with protected crossings and other pedestrian facilities on an area-wide level. This type of engineering and traffic control interventions requires some knowledge on pedestrians walking and crossing behaviour along entire trips. For example, the analysis of pedestrians crossing behaviour at a specific junction may provide useful insight for improving the design and operational elements of that junction. However, the analysis of pedestrians behaviour along entire trips from e.g. a metro station may reveal a need for area-wide interventions for the improvement of pedestrians level-of-service.

The proposed pilot models include a limited number of explanatory variables; nevertheless, some interesting implications may still be identified. For instance, it is suggested that pedestrians tend to postpone road crossing, and therefore the introduction of pedestrian facilities only in the proximity of a metro station may not be sufficient, as more crossings may take place farther, in the wider metro station area. Moreover, a need for better crossing facilities at mid-block may be identified, in case there are mid-block crossing attractors (e.g. large shops) in the area. Roadside parking management may also affect mid-block crossing behaviour, given that the presence of parked vehicles was found to encourage mid-block crossings. The examination of more explanatory variables in the next stages of the research will certainly reveal additional factors affecting crossing behaviour and safety along urban trips, and additional related implications concerning the design of both junctions and mid-block crossing locations at an area-wide level.

Further research

This paper presents a theoretical framework for modeling pedestrians crossing decisions along a trip in urban areas. A topological consideration of pedestrian trips was first outlined, allowing to identify specific scenarios of crossing behavior and to limit the analysis to particular types of crossings. The proposed scenarios are based on simple yet meaningful considerations, including definition and decomposition of different trip shapes in the form of topological graphs, origin / destination configurations and classification of crossings, which form an overall framework for approaching a complex problem in a systematic way. Furthermore, most of these considerations have been tested and validated in previous research (Lassarre et al 2007; Yannis et al. 2007).

On the basis of the scenarios identified, specific model formulations from the family of discrete choice models are proposed for analyzing pedestrians primary crossing decisions along a trip, under the assumption of either a hierarchical or a sequential decision making process. For the estimation of the proposed models, a particular data collection process is required, namely a field survey for observing and recording numerous characteristics of real pedestrian trips. In total, more than forty variables on trip, pedestrian, road, traffic and crossings characteristics are to be collected.

A pilot implementation of such a survey has already taken place, confirming the efficiency of such a data collection scheme, and preliminary modeling results on the basis of the data collected are promising, indicating that both approaches (hierarchical and sequential) are conceptually meaningful and computationally feasible. Nevertheless, a much more extensive survey, including several hundreds observations would be required and is in progress for obtaining interpretable results as per the explanatory effects sought, especially since the techniques considered in this research are quite demanding.

Further improvement in the proposed framework includes the consideration of more complex trips. In the pilot implementation, these were handled as combinations of the basis graph shapes presented above; however, more than one combination may be possible in each case and therefore extensive testing is required for identifying the most meaningful combinations. It is noted, however, that preliminary data collection confirmed that the vast majority of pedestrian trips conform to the four basic graph shapes presented in this research.

A final note concerns the type of models opted for in the proposed framework. In the two approaches considered, multinomial, nested and cross-nested structures are examined, whereas the only random effect considered concerns heterogeneity. Another useful related option would involve the use of mixed models, in which the assumption of the extreme value distribution of the random utility is relaxed, by allowing normally distributed error components. These models may provide even more flexibility in considering complex correlations among the observations.

References

Antonini G., Bierlaire M., Weber, M (2006). Discrete choice models of pedestrian walking behavior. *Transportation Research Part B* 40, 667-687.

Baltes M., Chu X. (2002). Pedestrian level of service for mid-block street crossings. *Transportation Research Record 1818*, 125-133.

Ben-Akiva, M., Lerman, S.R. (1985). *Discrete Choice Analysis: Theory and Applications to Travel Demand*. The MIT Press, Cambridge Massachusetts, London England.

Bierlaire, M. (2003). BIOGEME: A free package for the estimation of discrete choice models , Proceedings of the 3rd Swiss Transportation Research Conference, Ascona, Switzerland.

Bierlaire, M. (2006). A theoretical analysis of the cross-nested logit model. *Annals of Operations Research* 144 (1), 287-300.

Bierlaire, M. (2008). An introduction to BIOGEME Version 1.7. Available on-line at: biogeme.epfl.ch

Bierlaire, M., Antonini, G. and Weber, M. (2003). Behavioral dynamics for pedestrians, in K. Axhausen (ed) *Moving through nets: the physical and social dimensions of travel*, Elsevier.

Carsten, O.M.J., Sherborne, D.J., Rothengatter, J.A. (1998). Intelligent traffic signals for pedestrians: evaluation of trials in three countries. *Transportation Research Part C* 6.

Chu, X., Guttenplan, M., Baltes, M. (2004). Why people cross where they do - The role of street environment. *Transportation Research Record 1878,* 3-10.

Daly, A., Bierlaire M. (2004). A general and Operational representation of the generalized extreme value models. *Transportation Research part B* 40 (4), 285-305.

Das S., Manski C.F., Manuszak M. (2005). Walk or Wait? An Empirical Analysis of Street Crossing Decisions. *Journal of Applied Econometrics*. 20 (4), 445-577.

Fitzpatrick, K., Ullman, B., Trout, N. (2004). On-Street Pedestrian Surveys of Pedestrian Crossing Treatments. In the Proceedings of the *TRB 2004 Annual Meeting*, CD-ROM, Transportation Research Board, Washington DC.

Guggenheimer H. (1977). The Jordan curve theorem and an unpublished manuscript by max dehn. Archive for History of Exact Sciences 17 (2), 193-200.

Hakkert, S., Gitelman, V., Ben-Shabat, E. (2002). An evaluation of crosswalk warning systems: effects on pedestrian and vehicle behavior. *Transportation Research Part F* 5, 275-292.

Hamed M.M. (2001). Analysis of pedestrians' behavior at pedestrian crossings. *Safety Science* 38, 63-82.

Hine J. (1996). Pedestrian travel experiences: Assessing the impact of traffic on behavior and perceptions of safety using an in-depth interview technique. *Journal of Transport Geography* 4 (3), 179-199.

Honore B.E., Kyriazidou E. (2000). Panel Data Discrete Choice Models with Lagged Dependent Variables. *Econometrica* 68 (4), 839-874.

Hoogendoorn S.P. Pedestrian flow modeling by adaptive control. *Transportation Research Record 1878,* 95-103.

Hubbard S.M.L.,Bullock D.M., Mannering F.L. (2009). Right Turns on Green and Pedestrian Level of Service: Statistical Assessment. *Journal of Transportation Engineering* 135 (4), 153-159.

Ishaque, M. M., Noland, R. B. (2007). Behavioral issues in pedestrian speed choice and street crossing behavior: A review. Transport Reviews, 1-25.

Keegan O., O'Mahony M. (2003). Modifying pedestrian behavior. *Transportation Research Part A* 37, 889-901.

Lassarre, S., Papadimitriou E., Golias, J., Yannis, G. (2007). Measuring accident risk exposure for pedestrians in different micro-environments. *Accident Analysis and Prevention* 39 (6), 1226-1238.

Moore G.H. (2008). The emergence of open sets, closed sets, and limit points in analysis and topology. *Historia Mathematica* 35, 220-241.

Nee, J., Hallenberg, M. (2003). *A motorist and pedestrian behavioral analysis relating to pedestrian safety improvements*. Final Report, Research Project T1803, Task 16, Washington State Transportation Commission, USDOT-Federal Highway Administration.

Papadimitriou, E., Yannis G., Golias J. (2009). A critical assessment of pedestrian behavior models. Article in press. *Transportation Research Part F*, doi:10.1016/j.trf.2008.12.004.

Sisiopiku, V., Akin, D. (2003). Pedestrian behaviors at and perception towards various pedestrian facilities: an examination based on observation and survey data. *Transportation Research Part F* 6 (4), 249-274.

Van Houten R., Malenfant L. (1992). The influence of signs prompting motorists to yield before marked crosswalks on motor vehicle-pedestrian conflicts at crosswalks with flashing amber. *Accident Analysis & Prevention* 24 (3), 217-225.

Wooldridge, J.M. Simple solutions to the initial conditions problems in dynamic, non linear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics* Vol. 20, 2005, pp. 39-54.

Yang J., Deng W., Wang J., Li Q., Wang Z. (2006). Modeling pedestrians road crossing behavior in traffic system micro-simulation in China. *Transportation Research Part A* 40, 280-290.

Yannis G., Golias J., Papadimitriou E. (2007). Modeling crossing behavior and accident risk of pedestrians. *ASCE Journal of Transportation Engineering* 133 (11), 634-644.

Figure 1. a (top panel). Pedestrian trips as topological graphs, b (bottom panel). Alternative paths for a given configuration, primary and secondary crossings.

Figure 2. The Jordan curve theorem of topology

Figure 3. Basic graphs of pedestrian trips

Figure 4. Pairwise opposite crossings in a gamma-shaped trip graph with origin and destination in the interior set (not applicable scenario).

Figure 5. Primary crossing scenarios for gamma-shaped trip graphs and related choice sets

Figure 6. Nested Logit Model for the location of a primary crossing

Figure 7. Cross Nested Logit Model for the location of a primary crossing

Figure 8. Sequential Multinomial Logit Models for the location of a primary crossing

Table 1. Pedestrian primary crossing scenarios in relation to trip topology

Graph type	Origin set	Destination set	Primary crossings	Graph choice set
Linear	same		N=0	
	different		N=1	all
Gamma	Interior	Interior	N=0	
	Exterior	Exterior	N=0	
			N=2	N_1: link 1 N_2: link 2
	Interior	Exterior	N=1	N_1: link 2
	Exterior	Interior	N=1	N_1: link 1
Pi	Interior	Interior	N=0	
	Exterior	Exterior	N=0	
			N=2	N_1: link 1 N_2: link 3
	Interior	Exterior	N=1	N_1: links 2-3
	Exterior	Interior	N=1	N_1: links 1-2
Sigma	Interior	Interior	N=0	
			N=2	N_1: link 2
				N_1: link 2
	Exterior	Exterior	N=0	
			N=2	N_1: link 1 N_2: link 2
	Interior	Exterior	N=1	N_1: link 2
	Exterior	Interior	N=1	N_1: links 1-3
			N=3	N_1: link 1 N_2: link 2
				N_3: link 3

Table 2. Variables to be collected in the field survey

Variable	**Description**
Trip/person ID	Unique identification for each trip/person
Station	The origin station
Date	Date
Time	Time of the day
R_Descri	Route description by street names
R_Length	Total length of the trip (m)
R_time	Total duration of the trip (min)
R_shape	Shape of the trip graph (Linear, Gamma, Pi, Sigma)
R_Odtype	Origin/destination configuration (interior, exterior)
R_sc	The scenario to which this trip belongs
R_nbend	Total number of bends of the trip graph
R_Ctotal	Total number of crossings of the trip
R_Cptotal	Total number of primary crossings of the trip
R_Ltotal	Total number of road links of the trip
L_sn	Link serial number
L_choice	Road link choice indicator, (the link was chosen for primary crossing)
L_lanes	The number of lanes of the road link
L_oneway	The road link is one way
L_shoulder	The road link has a wide / narrow shoulder
L_median	The road link has a median
L_park	There is roadside parking on the link
L_guardrails	There are roadside guardrails on the link
L_traffic	The traffic volume on the link (low, high, congestion)
L_secend	A secondary crossing takes place at the end of the link
L_bendend	A bend of the graph takes place at the end of the link
Cp_sn	Primary crossing serial number
Cp_choice	Choice indicator for junction or mid-block
Cp_MBcrossw	There is a marked crosswalk at mid-block
Cp_Jcrossw	There is a marked crosswalk at junction
Cp_Jsignal	There is a traffic signal at junction
Cp_gap	The traffic gap accepted for the primary crossing
Cp_signal	The traffic signal display during the primary crossing
Cp_length	The distance from the trip origin where the primary crossing takes place
Cp_time	The time from the beginning of the trip where the primary crossing takes place
Cp_shelter	There is shade/shelter across the road where the crossing takes place
Cp_attract	There is an attractor (e.g. shop) across the road where the crossing takes place
Cp_acrossOD	The primary crossing takes place across the origin / the destination
Cp_following	The pedestrian is following another pedestrian making the same primary crossing

Cs_gap	The traffic gap accepted for the secondary crossing
Cs_signal	The traffic signal display for the secondary crossing
I_gender	The gender of the pedestrian
I_age	The age group of the pedestrian
I_hurry	The pedestrian appears to be in a hurry
I_carry	The pedestrian is carrying things
I_alone	The pedestrian is accompanied
I_Dfixed	The destination of the pedestrian is fixed

Table 3. Pilot implementation of a binary choice model (for choice sets including only one road link)

Utility parameters	Value	Robust Std. error	Robust t-test	p-value	
ASC0	0.00	fixed			
ASC1	2.83	1.39	2.03	0.04	
B_attract	-2.04	0.86	-2.38	0.02	
B_park	-1.25	0.61	-2.05	0.04	
B_crossw	0.68	0.54	1.26	0.21	*
B_gender	0.98	0.46	2.14	0.03	
B_loglength	-0.72	0.28	-2.55	0.01	
B_lsn1	0.00	fixed			
B_lsn23	0.84	0.56	1.51	0.13	*
B_lsn45	1.38	0.84	1.64	0.10	*

Utility functions

0: Midblock	ASC0 * one
1: Junction	ASC1 * one + B_gender * I_gender + B_park * L_park + B_crossw * J_crossw + B_attract * L_attract + B_loglength * R_loglength + B_lsn1 * L_sn1 + B_lsn23 * L_sn23 + B_lsn45 * L_sn45

Estimated parameters: 8
Observations: 127
Init log-likelihood: -88.03
Final log-likelihood: -59.42
Likelihood ratio test: 57.21
Rho-square: 0.33
Diagnostic: Convergence reached
Iterations: 7

Note: ASC refers to alternative-specific constant, one=1, * indicates a non significant effect

Table 4. Pilot implementation of sequential (multinomial) choice models

Utility parameters	**Value**	**Robust Std. error**	**Robust t-test**	**p-value**	
ASC0	3.74	1.71	2.19	0.03	
ASC1	3.16	1.69	1.88	0.06	*
ASC2	0.00	fixed			
B0_crossw	0.40	0.51	0.80	0.43	*
B1_crossw	1.38	0.52	2.68	0.01	
B2_crossw	0.00	fixed			
B_gender	0.55	0.37	1.50	0.13	*
B_loglength	0.90	0.31	2.87	0.00	
B_notprevious	0.88	0.41	2.14	0.03	
B_notprevious2	0.38	0.44	0.86	0.39	*
SIGMA	0.09	0.14	0.68	0.50	*
ZERO	0.00	fixed			

Utility functions

0: Mid-block	ASC0 * one + B_gender * I_gender + B0_crossw * J_crossw + B_notprevious * notprevious + B_notprevious2 * notprevious2 + ZERO [SIGMA] * one
1: Junction	ASC1 * one + B_gender * I_gender + B1_crossw * J_crossw + B_notprevious * notprevious + B_notprevious2 * notprevious2 + ZERO [SIGMA] * one
2: No crossing	ASC2 * one + B_loglength * R_loglength + B2_crossw * J_crossw

Estimated parameters:	9
Observations:	142
Individuals:	68
Init log-likelihood:	-156.00
Final log-likelihood:	-140.22
Likelihood ratio test:	31.57
Rho-square:	0.10
Diagnostic:	Convergence reached
Iterations:	122
Number of draws:	1000

Note: ASC refers to alternative-specific constant, one=1, ZERO[SIGMA] is a random panel effect ~N(0, σ^2), * indicates a non significant effect

Table 5. Pilot implementation of hierarchical (cross-nested) choice model

Utility parameters	**Value**	**Robust Std err**	**Robust t-test**	**p-value**	
ASC1	0.00	fixed			
ASC2	-2.03	3.35	-0.61	0.54	*
ASC3	-2.24	3.21	-0.70	0.48	*
ASC4	-5.33	3.34	-1.60	0.11	*
B1_loglength	1.00	fixed			
B2_loglength	1.72	0.62	2.79	0.01	
B3_loglength	1.75	0.60	2.91	0.00	
B4_loglength	2.29	0.60	3.84	0.00	
B_crossw	1.03	0.70	1.47	0.14	*
B_gender	-1.96	0.68	-2.90	0.00	

Model parameters	**Value**	**Robust Std err**	**Robust t-test**	**p-value**	
NEST1	1.00	fixed			
NEST2	3.10	9.37	0.33	0.74	*
NEST1_Alt1	1.00	fixed			
NEST1_Alt2	1.00	fixed			
NEST1_Alt3	0.50	fixed			
NEST1_Alt4	0.50	fixed			
NEST2_Alt1	0.00	fixed			
NEST2_Alt2	0.00	fixed			
NEST2_Alt3	0.50	fixed			
NEST2_Alt4	0.50	fixed			

Utility functions

1: Junction 1	ASC1 * one + B1_loglength * R_loglength + B_crossw * J_crossw
2: Mid-block 1	ASC2 * one + B2_loglength * R_loglength + B_gender * I_gender
3: Junction 2,3,4,5	ASC3 * one + B3_loglength * R_loglength + B_gender * I_gender
4: Mid-block 2,3,4,5	ASC4 * one + B4_loglength * R_loglength + B_gender * I_gender

Estimated parameters:	9
Observations:	75
Init log-likelihood:	-109.36
Final log-likelihood:	-95.03
Likelihood ratio test:	17.89
Rho-square:	0.09
Diagnostic:	Convergence reached
Iterations:	44

Note: ASC refers to alternative-specific constant, one=1, * indicates a non significant effect
NESTi_Altj refers to the membership factor of alternative (j) to nest (i)